

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE ARTES E LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

William Dubois

**O TESLLE E O TOEFL ITP: UMA ANÁLISE CRÍTICA DA TESTAGEM DE  
PROFICIÊNCIA EM LEITURA DE INGLÊS NA UFSM**

Santa Maria, RS

2021



**William Dubois**

**O TESLLE E O TOEFL ITP: UMA ANÁLISE CRÍTICA DA TESTAGEM DE  
PROFICIÊNCIA EM LEITURA DE INGLÊS NA UFSM**

Dissertação apresentada ao Programa de Pós-graduação em Letras, área de concentração em Estudos Linguísticos, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do título de **Mestre em Letras**.

Orientadora: Prof<sup>a</sup>. Dra. Patrícia Marcuzzo

Santa Maria, RS  
2021

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Oliveira, William Dubois  
O TESLLE E O TOEFL ITP: UMA ANÁLISE CRÍTICA DA  
TESTAGEM DE PROFICIÊNCIA EM LEITURA DE INGLÊS NA UFSM /  
William Dubois Oliveira.- 2021.  
123 p.; 30 cm

Orientadora: Patricia Marcuzzo  
Dissertação (mestrado) - Universidade Federal de Santa  
Maria, Centro de Artes e Letras, Programa de Pós-Graduação  
em Letras, RS, 2021

1. Testes 2. Proficiência 3. Leitura 4. TESLLE 5.  
TOEFL ITP I. Marcuzzo, Patricia II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Fatta CRB 10/1728.

Declaro, WILLIAM DUBOIS OLIVEIRA, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

**William Dubois**

**O TESLLE E O TOEFL ITP: UMA ANÁLISE CRÍTICA DA TESTAGEM DE  
PROFICIÊNCIA EM LEITURA DE INGLÊS NA UFSM**

Dissertação apresentada ao Programa de Pós-graduação em Letras, área de concentração em Estudos Linguísticos, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do título de **Mestre em Letras**.

**Aprovado em 07 de maio de 2021:**

---

**Patrícia Marcuzzo, Dra. (UFSM) - Videoconferência**  
(Presidente/Orientadora)

---

**Simone Sarmento, Dra. (UFRGS) - Videoconferência**

---

**Luciane Kirchhof Ticks, Dra. (UFSM) - Videoconferência**

Santa Maria, RS  
2021

**NUP:** 23081.038719/2021-07

**Prioridade:** Normal

**Homologação de ata de banca de defesa de pós-graduação**

134.332 - Bancas examinadoras: indicação e atuação

**COMPONENTE**

<b>Ordem</b>	<b>Descrição</b>	<b>Nome do arquivo</b>
2	Folha de Aprovação	Folha de aprovação.pdf

**Assinaturas**

**13/05/2021 11:49:16**

SIMONE SARMENTO (Pessoa Física)

Usuário Externo (458.\*\*\*.\*\*\*-\*\*)

**13/05/2021 13:03:58**

PATRICIA MARCUZZO (PROFESSOR DO MAGISTÉRIO SUPERIOR)

08.37.00.00.0.0 - DEPARTAMENTO DE LETRAS ESTRANGEIRAS E MODERNAS - DLTE

**13/05/2021 13:44:32**

LUCIANE KIRCHHOF TICKS (PROFESSOR DO MAGISTÉRIO SUPERIOR)

08.37.00.00.0.0 - DEPARTAMENTO DE LETRAS ESTRANGEIRAS E MODERNAS - DLTE

**Código Verificador:** 650046

**Código CRC:** 489b30af

**Consulte em:** <https://portal.ufsm.br/documentos/publico/autenticacao/assinaturas.html>



## RESUMO

### O TESLLE E O TOEFL ITP: UMA ANÁLISE CRÍTICA DA TESTAGEM DE PROFICIÊNCIA EM LEITURA DE INGLÊS NA UFSM

AUTOR: William Dubois  
ORIENTADORA: Patrícia Marcuzzo

Para a permanência em programas de pós-graduação, no Brasil, os pós-graduandos precisam comprovar suficiência ou proficiência linguística em, pelo menos, uma língua estrangeira moderna. Na Universidade Federal de Santa Maria (UFSM), essa comprovação é obtida principalmente por meio dos resultados de testes como o Teste de Suficiência em Leitura em Língua Estrangeira (TESLLE) e a seção de leitura do TOEFL-ITP (*Test of English as a Foreign Language - Institutional Testing Program*). Dessa forma, fundamentando-se na Análise Crítica de Gênero (ACG), o presente estudo relata resultados de investigações acerca desses testes enquanto práticas sociais de testagem de inglês para fins acadêmicos (*English for Academic Purposes - EAP*), a partir de documentos e amostras oficiais destes. Na análise contextual, foram identificadas as descrições oficiais, as características estruturais desses testes enquanto gênero, as condições de aplicação, as médias de pontuação, os objetivos e os públicos-alvo. Na análise textual, os textos-base foram analisados com relação ao gênero, aos conteúdos e às características estruturais. A análise das questões enfatizou a identificação das estruturas, dos tipos, dos enfoques e dos estratos que elas envolvem. A partir da relação indissociável entre texto e contexto, levantamos discussões sobre as concepções de proficiência, leitura e (multi)letramento(s) que subjazem aos testes, bem como sobre confiabilidade, validade e efeito retroativo. A concepção de proficiência em leitura de EAP do TESLLE diz respeito compreensão dos aspectos da léxico-gramática, semântica e pragmática, do gênero e do(s) discurso(s) de um texto autêntico, ao passo que, para a seção de leitura do TOEFL ITP, essa proficiência significa entender com imediatez e objetividade os aspectos semânticos e pragmáticos de textos breves e descontextualizados. Espera-se que este estudo contribua para a área de Linguística Aplicada e que os resultados ajudem a promover uma reflexão a respeito da importância e do papel social dessas práticas de testagem de leitura, que poderão contribuir para que esses testes sejam instrumentos benéficos para a sociedade.

**Palavras-chave:** Testes. Proficiência. Leitura. TESLLE. TOEFL ITP.



## ABSTRACT

### THE TESLLE AND THE TOEFL ITP: A CRITICAL ANALYSIS OF THE UFSM ENGLISH READING PROFICIENCY TESTING

AUTHOR: William Dubois  
ADVISOR: Patrícia Marcuzzo

For the permanence in postgraduate programs in Brazil, students must prove sufficiency or proficiency in, at least, one modern foreign language. At the Federal University of Santa Maria (UFSM), this proof is mainly obtained through the results of tests such as the *Teste de Suficiência em Leitura em Língua Estrangeira* (TESLLE) and the reading section of the TOEFL-ITP (*Test of English as a Foreign Language - Institutional Testing Program*). In this way, based on the Critical Genre Analysis (CGA), the present study reports results of a research on these tests as social practices of testing of English for Academic Purposes (EAP), from official documents and samples of these tests. In the contextual analysis, the official descriptions, the structural characteristics of these tests as a genre, the conditions of their application, the score averages, the objectives and the target-audiences were identified. In the textual analysis, the base-texts were analyzed in terms of their genres, contents and structural characteristics. The analysis of the questions focused on the identification of the structures, types, foci and the language strata they encompass. From the inseparable relationship between text and context, we discuss the concepts of proficiency, reading and (multi)literacy that underly these tests, as well as reliability, validity and washback effect. The concept of EAP reading proficiency of the TESLLE regards the aspects of the lexicogrammar, the semantics and pragmatics, the genre and the discourse(s) of an authentic text, whereas, for the reading section of the TOEFL ITP, this proficiency means to quickly and objectively understand the semantic and pragmatic aspects of short and decontextualized texts. This study is expected to contribute to the area of Applied Linguistics and that its results help promoting a reflection on the social role and importance of these social practices of language testing, which may contribute to making these tests beneficial instruments for the society.

**Keywords:** Reading. Proficiency. Tests. TESLLE. TOEFL ITP.



## LISTA DE ILUSTRAÇÕES

Figura 1 - Concepção tridimensional de Análise Crítica do Discurso	17
Figura 2 - Articulação das concepções de discurso como texto, interação e contexto e de linguagem como sistema sociossemiótico	34
Figura 3 - Aportes teórico-metodológicos da Análise Crítica de Gênero	36
Figura 4 - O TESLLE de Língua Inglesa	50
Figura 5 - A seção de leitura do TOEFL ITP	52
Figura 6 - Texto-base do TESLLE de Língua Inglesa	59
Figura 7 - Textos-base da seção de leitura do TOEFL ITP	61
Figura 8 - Estrutura das questões do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP	65
Figura 9 - Questões da seção de leitura do TOEFL ITP com enfoque em equivalência semântico-lexical	72
Figura 10 - Equivalência do número de acertos em pontos na seção de leitura do TOEFL ITP	82



## LISTA DE QUADROS

Quadro 1 - Tipos de testes em relação a seus propósitos	20
Quadro 2 - Relação entre as competências, estratificação da linguagem e enfoques das questões	32
Quadro 3 - O <i>corpus</i> deste estudo	46
Quadro 4 – Organização retórica do gênero teste de leitura de EAP	54
Quadro 5 – Descrições dos textos-base do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP.	58
Quadro 6 - Descrições das questões do TESLLE e da seção de leitura do TOEFL ITP	64
Quadro 7 - Estrutura das questões do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP	66
Quadro 8 - Tipos de questão do TESLLE de Língua Inglesa	68
Quadro 9 - Tipos de questão da seção de leitura do TOEFL ITP	69
Quadro 10 - Enfoques de questão do TESLLE.	70
Quadro 11 - Enfoques de questão da seção de leitura do TOEFL ITP	73
Quadro 12 - Estratos e competências envolvidas nas questões do TESLLE	74
Quadro 13 - Questões com referência explícita a fragmentos do texto-base	75
Quadro 14 - Aspectos contextuais dos testes	76
Quadro 15 - <i>Scores</i> do TESLLE	81
Quadro 16 - Descrições de proficiência de acordo com o nível QCER e a pontuação na seção de leitura do TOEFL ITP	81
Quadro 17 - Síntese dos resultados	84



## LISTA DE ABREVIATURAS E SIGLAS

ACD	Análise Crítica de Discurso
ACG	Análise Crítica de Gênero
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CEFR	<i>Common European Framework of Reference for Languages</i>
CSF	Ciência sem Fronteiras
DLEM	Departamento de Letras Estrangeiras Modernas
EAP	<i>English for Academic Purposes</i>
ENEM	Exame Nacional do Ensino Médio
ETS	<i>Educational Testing Service</i>
GRE	<i>Graduate Record Examination</i>
GSF	Gramática Sistemico-Funcional
iBT	<i>Internet Based Test</i>
IES	Instituições de Ensino Superior
ISF	Idiomas sem Fronteiras/Inglês sem Fronteiras
ITP	<i>Institutional Testing Program</i>
LABLER	Laboratório de Pesquisa e Ensino de Leitura e Redação
LINC	Línguas no Campus
LSF	Linguística Sistemico-Funcional
MEC	Ministério da Educação
PBT	<i>Paper Based Test</i>
PPGL	Programa de Pós-graduação em Letras
PRAE	Pró-reitoria de Assuntos Estudantis
PROGRAD	Pró-reitoria de Graduação
PRPGP	Pró-reitoria de Pós-graduação e Pesquisa
QCER	Quadro Comum Europeu de Referência
SCIELO	<i>Scientific Electronic Library Online</i>
TESLLE	Teste de Suficiência em Leitura em Língua Estrangeira
TOEFL	<i>Test of English as a Foreign Language</i>
TOEIC	<i>Test of English for International Communication</i>
UFRGS	Universidade Federal do Rio Grande do Sul
UFSC	Universidade Federal de Santa Catarina
UFSM	Universidade Federal de Santa Maria



## SUMÁRIO

<b>CAPÍTULO 1 - INTRODUÇÃO</b>	<b>12</b>
1.1 OBJETIVOS	15
1.2 A ORGANIZAÇÃO DESTE TRABALHO	16
<b>CAPÍTULO 2 - REVISÃO DA LITERATURA</b>	<b>17</b>
2.1 A PRÁTICA SOCIAL DA TESTAGEM DE PROFICIÊNCIA DE LEITURA DE EAP	17
2.2 OS TESTES DE LEITURA DE EAP	19
<b>2.2.1 O TESLLE (Teste de Suficiência em Leitura em Língua Estrangeira)</b>	<b>23</b>
<b>2.2.2 O TOEFL ITP (<i>Test of English as a Foreign Language - Institutional Testing Program</i>)</b>	<b>26</b>
2.3 ENFOQUES E TIPOS DE QUESTÃO	31
2.4 A ANÁLISE CRÍTICA DE GÊNERO	33
2.5 LINGUAGEM, LEITURA E (MULTI)LETRAMENTO(S)	38
2.6 CONFIABILIDADE, VALIDADE E EFEITO RETROATIVO	40
<b>CAPÍTULO 3 - METODOLOGIA</b>	<b>45</b>
3.1 O <i>CORPUS</i> : CRITÉRIOS DE SELEÇÃO E DESCRIÇÃO DA AMOSTRA	45
3.2 OS PROCEDIMENTOS DE ANÁLISE	47
<b>3.2.1 Análise contextual</b>	<b>47</b>
<b>3.2.2 Análise textual</b>	<b>48</b>
<b>CAPÍTULO 4 - RESULTADOS E DISCUSSÃO</b>	<b>50</b>
4.1 A ESTRUTURA DOS TESTES	50
<b>4.1.1 Os textos-base</b>	<b>57</b>
<b>4.1.2 As questões</b>	<b>63</b>
4.1.2.1 <i>Aspectos contextuais das questões</i>	63
4.1.2.2 <i>As estruturas das questões</i>	64
4.1.2.3 <i>Tipos de questão</i>	66
4.1.2.4 <i>Enfoques de questão</i>	70
4.1.2.5 <i>Estratos e competências das questões</i>	73
4.2 O TESLLE E O TOEFL ITP NA PRÁTICA SOCIAL	76
4.3 SISTEMATIZAÇÃO E INTERPRETAÇÃO DOS RESULTADOS	84
<b>CAPÍTULO 5 - CONSIDERAÇÕES FINAIS</b>	<b>93</b>
<b>REFERÊNCIAS</b>	<b>95</b>
<b>APÊNDICE A - AMOSTRA DO TESLLE DE LÍNGUA INGLESA</b>	<b>102</b>



## CAPÍTULO 1 - INTRODUÇÃO

É de amplo consenso que ler publicações de cunho científico em língua estrangeira/adicional, especialmente inglês, é uma necessidade de pesquisadores das mais variadas áreas do conhecimento, uma vez que o inglês tem sido visto como a “língua franca da ciência” (e.g., JENKINS, 2014; FINARDI, 2018). Atualmente, grande parte da produção científica formal é divulgada e publicada substancialmente em textos escritos em inglês, como, por exemplo, artigos acadêmicos. Por meio da leitura dessas publicações, esses pesquisadores de todo o mundo visam a se inteirar das pesquisas realizadas em diferentes contextos para então desenvolverem e publicarem suas próprias pesquisas.

Como um reflexo disso, nos últimos anos, uma parcela importante das instituições de ensino no Brasil têm incluído a habilidade de leitura em inglês para fins acadêmicos (à qual doravante será referida pela sigla EAP, em inglês, *English for Academic Purposes*) (SWALES, 1990; BHATIA, 1993) como um dos (pré) requisitos em diversos editais e resoluções, como, por exemplo, os que dizem respeito ao ingresso, à permanência ou à conclusão de cursos de pós-graduação. Pesquisadores em formação, no contexto dos programas de pós-graduação, precisam comprovar que conseguem ler textos em língua estrangeira/adicional para frequentar ou então se manter como alunos nesses cursos. Essa comprovação é realizada por meio de testes (chamados de testes de proficiência) que sejam aceitos de acordo com critérios estipulados pelos programas de pós-graduação e pelas instituições de ensino às quais estão vinculados.

Nesse contexto, testes de leitura de EAP têm sido amplamente desenvolvidos para a comprovação de proficiência nessa habilidade. Esses testes são utilizados e até mesmo comercializados para esse fim (WIELEWICKI, 1997). Atualmente, a relevância desses testes é expressiva no contexto acadêmico brasileiro. Dentre os motivos para isso, podem-se mencionar o número crescente de programas de pós-graduação no Brasil e, conseqüentemente, pesquisadores em formação (MEC, 2020), da adoção do inglês entre os periódicos SCIELO (*Scientific Electronic Library Online*) (PACKER, 2016), etc.

No entanto, ainda há poucos estudos acerca de testes, como apontam McNamara e Roever, “a testagem de língua têm um histórico longo como uma prática social, mas relativamente curto como uma prática institucional teoricamente fundamentada e auto-reflexiva” (2006, p. 1),<sup>1</sup> (2006, p. 1). Por exemplo, apesar do caráter internacional e da ampla popularidade dos testes dos testes TOEFL (*Test of English as a Foreign Language*) no meio acadêmico internacional, não foram encontrados estudos sobre estes e sua relação com a prática social da testagem de proficiência em leitura de EAP no contexto do Brasil. Nesse sentido, este estudo se concentra no contexto específico da Universidade Federal de Santa Maria (UFSM) e os dois principais testes que fazem parte dessa prática social na instituição: o TESLLE (Teste de Suficiência em Leitura em Língua Estrangeira) e o TOEFL ITP (*Institutional Testing Program*).

Na UFSM, os critérios de comprovação de proficiência linguística em inglês como língua estrangeira/adicional são estabelecidos por, pelo menos, três resoluções: i) a resolução n. 003/2010 (UFSM, 2010), que estabelece normas relativas ao TESLLE e aproveitamento de testes de outras instituições, como os testes TOEFL por exemplo; ii) a resolução n. 015/2014 (UFSM, 2014), que aprova o Regimento Geral da Pós-graduação *Stricto Sensu* e *Lato Sensu* da UFSM, cujo artigo 31 trata brevemente da comprovação de proficiência em língua estrangeira; e iii) a resolução n. 018/2018 (UFSM, 2018), que institui a Política Linguística da UFSM e estabelece um conjunto de princípios e ações estratégicas que visam a suprir as demandas linguísticas da universidade. Essas resoluções se enquadram nas orientações da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), fundação do Ministério da Educação (MEC) que promove a expansão e consolidação da pós-graduação *stricto sensu* (mestrado e doutorado) no Brasil<sup>2</sup>, e cujos programas de pós-graduação credenciados por esta devem orientar seus alunos para comprovarem a proficiência.

A resolução n. 003/10 estipula o público-alvo e a periodicidade mínima do TESLLE, além dos critérios para o aproveitamento de testes de outras instituições, como os testes TOEFL, por exemplo. Dentre os artigos, está atribuída ao Departamento de Letras Estrangeiras Modernas (DLEM), vinculado ao Centro de Artes e Letras (CAL), a

---

<sup>1</sup> *Language testing has a long history as a social practice but only a relatively short one as a theoretically founded and selfreflexive institutional practice.* Tradução nossa.

<sup>2</sup> Fonte: <https://uab.capes.gov.br/perguntas-frequentes> Acesso em: 23 de jan. 2021.

competência exclusiva de realizar os procedimentos (elaboração, aplicação e correção) de testagem<sup>3</sup> relativos ao TESLLE (UFSM, 2010).

Por sua vez, o artigo 31 da resolução n. 015/2014 estipula que

Art. 31 Os discentes de pós-graduação em nível de Mestrado e Doutorado deverão comprovar suficiência em, no mínimo, uma língua estrangeira, conforme definido no regulamento do programa de pós-graduação.

§ 1º Uma vez homologada pelo colegiado do programa a comprovação da suficiência em língua(s) estrangeira(s), constará no histórico escolar do discente, com a expressão "Aprovado" ou "Reprovado".

§ 2º Os discentes poderão cumprir esse requisito de acordo com as opções e regulamentações definidas em legislação vigente da UFSM (UFSM, 2014).

Com essa resolução, determina-se que a aprovação em um teste de língua estrangeira se faz necessária para a permanência de alunos de pós-graduação em seus respectivos cursos e deverá constar no histórico escolar desses alunos. Dessa forma, a UFSM reafirma a importância de que se saiba ler suficientemente bem em pelo menos uma língua estrangeira no contexto acadêmico e científico do qual seus pós-graduandos fazem parte.

Por fim, a resolução n. 018/2018 (UFSM, 2018), que institui a Política Linguística da UFSM, foi pensada a partir de uma exigência do programa Idiomas Sem Fronteiras (IsF), e proposta por uma comissão integrada por sete representantes de diferentes áreas de letras (alemão, espanhol, francês, inglês, português para estrangeiros, português e literatura), além de representantes do Departamento de Educação Especial da área de libras, de três pró-reitorias e uma coordenação da universidade (da Pró-reitoria de Graduação (PROGRAD), da Pró-reitoria de Pós-graduação e Pesquisa (PRPGP), da Pró-reitoria de Assuntos Estudantis (PRAE), e da Coordenação de Educação Básica, Técnica e Tecnológica da UFSM. A implementação dessa política preocupa-se principalmente com a internacionalização e a difusão internacional das produções intelectuais, científicas, artísticas e culturais da UFSM. Para isso, suas diretrizes tratam do acesso democrático ao ensino de línguas, o respeito à diversidade linguística, a

---

<sup>3</sup> Neste trabalho, utilizamos o termo “testagem” ou “processo de testagem” para referir à prática social de testar algo, que engloba diversos outros processos e gêneros. Já o termo “teste” é utilizado para referirmo-nos ao gênero, que é o principal instrumento do processo de testagem.

formação crítica e cidadã do aprendiz de línguas, e garantir o letramento acadêmico e a difusão internacional das produções desenvolvidas pela comunidade da UFSM.

Dessa forma, em linhas gerais, o presente estudo busca investigar como a prática social da comprovação de proficiência em leitura de EAP se organiza frente as necessidades dos alunos de pós-graduação da UFSM. Esta pesquisa se justifica dada a emergente necessidade de que sejam analisados esses processos, especialmente em contextos específicos.

Além disso, o presente estudo foi motivado pela formação e atuação profissional do autor, que é professor licenciado em letras inglês e literaturas de língua inglesa pela UFSM. Além de trabalhar em cursos livres, também atuou no projeto Línguas no Campus (LINC) entre 2016 e 2017, vinculado ao Laboratório de Pesquisa e Ensino de Leitura e Redação (LabLeR), e posteriormente no programa Idiomas Sem Fronteiras (IsF), entre 2017 e 2019, ministrando cursos e elaborando materiais didáticos para alunos que prestaram os testes TESLLE e TOEFL ITP.

## 1.1 OBJETIVOS

O presente estudo tem por objetivo geral investigar a prática de testagem de proficiência em leitura de inglês para fins acadêmicos no âmbito acadêmico-institucional da UFSM, por meio da análise de dois testes amplamente conhecidos na instituição, a saber, o TESLLE e o TOEFL ITP. O gênero desses testes, que chamamos de teste de leitura de EAP, é a peça-chave para que se entenda essa prática. Para atingir o objetivo geral, delineamos três objetivos específicos:

1. analisar e descrever aspectos contextuais dos testes TESLLE e TOEFL ITP, identificando os objetivos sociocomunicativos, os públicos-alvo e as condições de aplicação destes;
2. analisar exemplares do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP com vistas a identificar e descrever como esses testes são estruturados textualmente; e, a partir disso
3. identificar e avaliar as concepções de linguagem, proficiência em leitura de EAP e letramento que subjazem a esses testes, a partir das análises e da interpretação

da relação indissociável entre texto e contexto, discutindo questões acerca da confiabilidade, da validade e do efeito retroativo desses testes nessa prática social.

Para isso, analisamos exemplares desses dois testes à luz da Análise Crítica de Gênero (ACG) (MEURER, 2002; 2003; 2004; MOTTA-ROTH, 2005; 2008), cujos pressupostos teórico-metodológicos enfatizam a necessidade e a importância de investigações correlacionadas de texto e contexto.

A partir dos resultados desta pesquisa, será possível entender, discutir e avaliar, com base em dados de análises científicas, tanto o gênero aqui analisado quanto os propósitos da prática social da testagem de proficiência em leitura de língua inglesa de EAP.

Ademais, este estudo está inserido na linha de pesquisa Linguagem no Contexto Social do Programa de Pós-Graduação em Letras (PPGL-UFSM), e foi contemplado com uma bolsa CAPES em março de 2020. Além disso, este trabalho está vinculado ao LabLeR na medida em que integra o projeto de pesquisa intitulado Análise Crítica de Gênero de testes de inglês como língua estrangeira (MARCUIZZO, 2018) (registro n°. 048796), que abarca diferentes testes de proficiência de leitura em língua inglesa (como o ENEM da área de Linguagens, Códigos e suas Tecnologias - Língua Inglesa, os vestibulares da Universidade Federal do Rio Grande do Sul - UFRGS e da Universidade Federal de Santa Catarina - UFSC, o TOEFL ITP e o TESLLE da UFSM).

## 1.2 A ORGANIZAÇÃO DESTE TRABALHO

Esta dissertação está organizada em cinco capítulos. No capítulo 1, de Introdução, a temática e os objetivos desta pesquisa são apresentados e contextualizados. No capítulo 2, de Revisão da Literatura, apresentam-se os conceitos e pressupostos teóricos essenciais para as análises realizadas neste trabalho. No capítulo 3, de Metodologia, são descritos o *corpus*, seus critérios de seleção e os procedimentos de análise realizados. No capítulo 4, de Resultados e Discussão, são descritas e interpretadas as descobertas oriundas das análises realizadas. Por fim, no capítulo 5, intitulado Considerações Preliminares, são indicadas as conclusões e as perspectivas futuras da pesquisa.

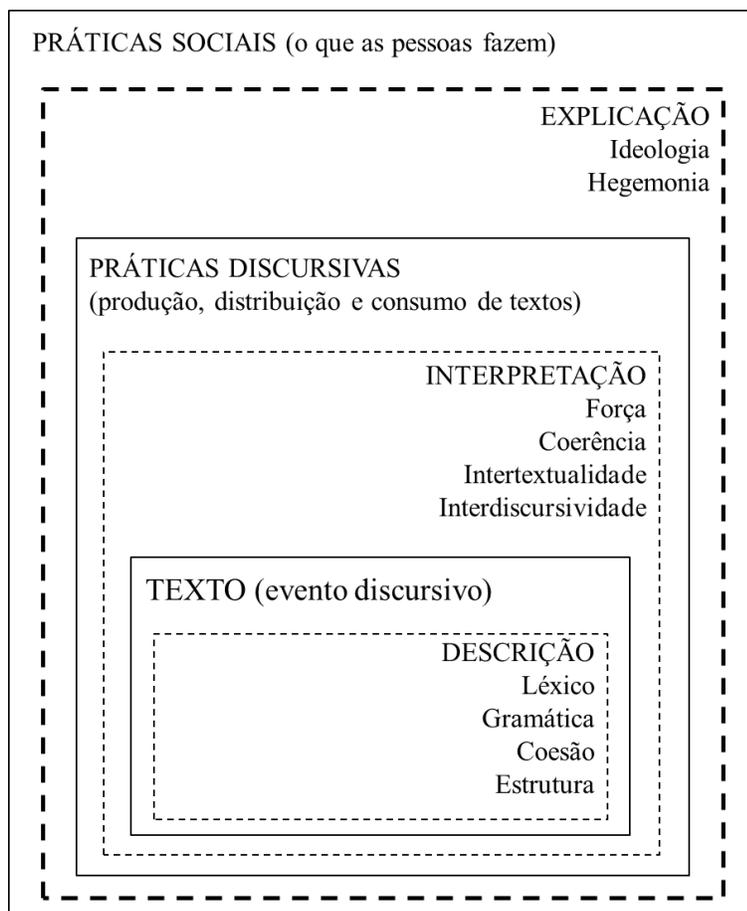
## **CAPÍTULO 2 - REVISÃO DA LITERATURA**

Neste capítulo, o enquadramento teórico-metodológico adotado na presente pesquisa é apresentado e discutido. Na seção 2.1 são revisados os pressupostos que embasam a concepção de prática social adotada pelo presente estudo. Na seção 2.2 é apresentado o conceito de gênero que nos leva à identificação dos testes de leitura de EAP. Os exemplares desse gênero investigados nesta pesquisa são descritos na seção 2.2.1, que trata do TESLLE, e na seção 2.2.2, que trata do TOEFL ITP. Posteriormente, as concepções que norteiam este estudo são apresentadas nas seções 2.3, que trata dos enfoques e tipos de questão; 2.4, que apresenta a Análise Crítica de Gênero; 2.5, que discute e relaciona concepções de linguagem, leitura e (multi)letramento(s); e 2.6, que trata dos conceitos de confiabilidade, validade e efeito retroativo.

### **2.1 A PRÁTICA SOCIAL DA TESTAGEM DE PROFICIÊNCIA DE LEITURA DE EAP**

Os conceitos de prática linguística, discursiva e social que norteiam este trabalho partem da Análise Crítica de Gênero (ACG). Por sua vez, essa perspectiva teórico-metodológica é fortemente influenciada pela Análise Crítica do Discurso (ACD), que inclui outros dois níveis cruciais para a análise de uma prática social: a prática discursiva e o texto, formando a concepção tridimensional do discurso proposta por Fairclough, na qual o texto está inserido na prática discursiva, que, por sua vez, está inserida na prática social (FAIRCLOUGH, 2001, p. 102), conforme apresenta a Figura 1.

Figura 1 - Concepção tridimensional de Análise Crítica do Discurso



Fonte: Meurer (2005, p. 95)

Nessa concepção, proposta por Fairclough (2001) e posteriormente elaborada por Meurer (2005), é estabelecido que o texto é a materialidade linguística na qual se manifesta a prática discursiva, que, por sua vez, é uma dimensão da prática social (apud MARCUZZO, 2011, p. 53). Portanto, para que se faça a análise do texto, Fairclough propõe que se analisem o vocabulário, a gramática, a coesão e a estrutura textual (Ibid., p. 103) e, para tanto, utiliza a Gramática Sistêmico-Funcional (GSF) proposta por Halliday (1994; 2004, com Matthiessen). A prática discursiva, por sua vez, diz respeito aos processos de produção, distribuição e consumo dos textos, que são o foco de análises dessa dimensão da linguagem. O terceiro foco dessa teoria tridimensional é a prática social, que é definida como “aquilo que as pessoas fazem, isto é, as atividades em que se engajam ao conduzir a vida social” (MEURER, 2004, p. 138). A análise dessa

dimensão visa a entender as posições e relações sociais de seus participantes, que é expressa por meio tanto do discurso quanto do texto (MARCUIZZO, 2011, p. 103).

Para que se analise gênero como uma prática discursiva socialmente situada, faz-se necessário analisar texto e contexto simultaneamente, pois, ainda de acordo com Fairclough, a relação entre estes é dialética, ou seja, uma é resultante da outra (FAIRCLOUGH, 2001, p. 91). Esse princípio está alinhado com a Gramática Sistêmico-Funcional (GSF), outra perspectiva que embasa a ACG, na qual entende-se que a relação entre texto e contexto é de interdependência (THOMPSON, 2004, p. 9), de modo que o texto cria e é criado por seu contexto (HALLIDAY, 1989, p. 47). Neste trabalho, é realizada a dimensão da prática social da testagem de proficiência em inglês como língua estrangeira/adicional.

## 2.2 OS TESTES DE LEITURA DE EAP

Em diversos contextos, testes são frequentemente entendidos como sinônimo de avaliações, exames ou provas. No entanto, a literatura da área mostra que existem diferenças entre testagem e avaliação (WIGGINGS, 1993; MARCHEZAN, 2005, dentre outros). Marchezan (2005, p. 27) aponta que uma avaliação é parte do ensino: é o processo do ensino e aprendizagem, enquanto um teste é algo independente do ensino, que tem objetivos próprios e não prevê contato entre o elaborador do teste e o examinando<sup>4</sup>, uma vez que é o produto do aprendizado.

Similarmente, Wiggings (1993, p. 4) afirma que, em um teste, há “uma distância entre o elaborador e o aluno”, considerando que o examinando não pode “adaptar a questão para seu estilo pessoal nem questionar o elaborador”, como é normalmente viável em avaliações (como provas e trabalhos escolares, por exemplo).

Isto posto, podemos afirmar que o objetivo principal de um teste é testar os conhecimentos e/ou as habilidades de seus examinandos acerca de determinado(s) assunto(s). Por ser independente de práticas de ensino, um teste não se propõe necessariamente a fornecer *feedback*. Portanto, é importante que se faça a distinção das

---

<sup>4</sup> Neste trabalho, chamamos de examinando/a(s) as pessoas que realizam um teste de leitura de EAP. Em português também é usado o termo “candidato”, a depender dos propósitos do teste em questão. Em inglês o termo mais comumente usado é *test taker*.

práticas de testagem, nas quais os testes se inserem, das práticas de ensino e aprendizado, ainda que estas sejam voltadas para o sucesso na realização de testes e/ou adapta-los como material didático ou instrumento de avaliação.

Nesse sentido, Hughes (1997) afirma que existem diversas controvérsias a respeito de testagens, que discutem a possibilidade de testes não serem o melhor tipo de avaliação para propósitos de ensino. Entretanto, eles são utilizados em muitas situações relacionadas com o ensino, como, por exemplo, no contexto de seleção para mobilidade acadêmica internacional. Ainda de acordo com Hughes (1997, p. 4), “informações sobre a habilidade linguística das pessoas são frequentemente muito úteis e algumas vezes necessárias”, e essas informações podem ser obtidas por meio de um processo de testagem.

É sabido que testes de leitura de EAP figuram entre as atividades mais comuns em âmbito acadêmico e institucional. Avelar (2015) classificou diversos testes de EAP em relação ao uso que se faz dos resultados, à função ou o propósito” (AVELAR, 2015, p. 34).

Quadro 1 - Tipos de testes em relação a seus objetivos

(continua)

Tipo de teste	Objetivo principal	Exemplos
Avaliações de rendimento ( <i>achievement tests</i> )	Avaliar o conteúdo trabalhado em sala de aula, ao longo de um processo de ensino	Avaliações de sala de aula, aplicadas por professores
Testes de proficiência ou suficiência ( <i>proficiency tests</i> )	Estimar o grau de proficiência em (determinadas habilidades de) língua adicional	FCE, CAE, CPE, TOEFL, IELTS, Celpe-Bras, etc.
Testes de colocação, nivelamento ou classificação ( <i>placement tests</i> )	Classificar aprendizes de acordo com níveis de uma determinada instituição	Exames aplicados em instituições de idiomas para classificar os aprendizes

## Quadro 1 - Tipos de testes em relação a seus objetivos

(continuação)

Exames de entrada ( <i>admission exams</i> )	Selecionar candidatos para uma determinada instituição ou um programa de ensino	Vestibulares, ENEM, etc.
---	--	--------------------------

Fonte: Elaborado pelo autor com base em Avelar (2015, p. 36).

Tomando a classificação de Avelar (2015) como parâmetro, é possível afirmar que o TESLLE, apesar de ser intitulado teste de suficiência (termo definido pela comissão do TESLLE, à época, para indicar que, com a aprovação no TESLLE, o aprovado seria capaz de ler suficientemente bem textos em inglês), se enquadra como um teste de proficiência, cujo propósito é aferir a habilidade de leitura em língua adicional/estrangeira dos alunos e servidores da UFSM. Por sua vez, o TOEFL ITP pode ser enquadrado em todos os tipos de teste identificados pela autora. Isso se deve, em parte, ao fato de que se trata de um teste internacional, que é utilizado para diversos fins. Em cursos livres de línguas no Brasil, por exemplo, o TOEFL ITP é utilizado tanto como uma avaliação de rendimento quanto como um teste de proficiência. No contexto do IsF, o TOEFL ITP costumava ser utilizado como um teste de nivelamento para os inúmeros cursos de língua inglesa que o programa oferecia. Além disso, o TOEFL ITP também se configura como um exame de entrada, por ser recorrentemente um pré-requisito para a realização de mobilidade acadêmica internacional, ou mesmo para a matrícula em universidades em diversos países.

No presente estudo, no entanto, analisamos o TESLLE de Língua Inglesa e a seção de leitura do TOEFL ITP como testes de proficiência em leitura de língua inglesa como língua adicional/estrangeira. Isso significa que nosso propósito é avaliar esses testes de forma a verificar de que forma e em que medida eles são bons indicadores de proficiência, sem necessariamente nos debruçarmos sobre os demais propósitos e usos que podem ser feitos dos testes e suas respectivas comprovações.

Assim sendo, neste estudo, o TESLLE de Língua Inglesa e a seção de leitura (*reading comprehension*) do TOEFL ITP são analisados e interpretados como

exemplares de um gênero discursivo multimodal, a que chamamos de teste de leitura de EAP<sup>5</sup>.

Além disso, ambos os exemplares são interpretados como testes (e não provas ou avaliações, como são popularmente chamados) por não serem instrumentos de ensino. Embora existam diversos cursos e materiais didáticos preparatórios para ambos os testes, o processo de testagem em que estão inseridos não implica o contato direto ou indireto entre os elaboradores dos testes e seus examinandos (MARCHEZAN, 2005, p. 27). Ainda, com exceção da nota disponibilizada por suas respectivas instituições responsáveis, testes não fornecem nenhum *feedback*, como é tipicamente o caso de provas e avaliações.

Esses testes são entendidos como gêneros pois são “usos da linguagem associados a atividades sociais” (MOTTA-ROTH, 2008, p. 350) e realizam uma “[...] atividade culturalmente pertinente, mediada pela linguagem num dado contexto de situação atravessado por discursos de ordens diversas” (MOTTA-ROTH, 2005, p. 181), sendo “um processo social organizado por estágios e orientado para objetivos” (MARTIN; MATTHIESSEN, 2012, p. 144). Esses usos da linguagem possuem uma estrutura esquemática que é organizada em movimentos retóricos, que são unidades discursivas que realizam, por meio de passos retóricos, uma função comunicativa em um texto oral ou escrito (SWALES, 2004, p. 228-229). Muito embora a seção de leitura seja parte do teste TOEFL ITP e possa ser entendida como um estágio ou movimento retórico<sup>6</sup> do teste TOEFL ITP, essa seção, assim como o TESLLE, também se caracteriza como um exemplar de gênero discursivo por si só, pois, nela há uma tipificação estrutural reconhecível em termos de pertinência linguística, eventos e participantes envolvidos na atividade comunicativa, além de um objetivo comunicativo específico.

Por fim, trata-se também de um gênero multimodal, uma vez que todos os textos são multimodais (KRESS, 1997). Muito embora as materialidades linguísticas que

---

<sup>5</sup> Cabe para distinguir que o termo leitura de inglês (como língua adicional/estrangeira e/ou para fins acadêmicos - EAP) abrange: i) testagens e testes de leitura *de* um ou mais textos-base em inglês, para responder questões escritas em português (ou na língua materna de determinado examinando), como o TESLLE; e ii) testagens e testes de leitura *em* inglês, nos quais o(s) texto(s)-base e as questões do teste são apresentadas integralmente em Língua Inglesa, como nos testes TOEFL.

<sup>6</sup> Martin e Matthiessen (2012) utilizam o termo estágios, enquanto Swales (2004) utiliza o termo movimentos retóricos. Neste trabalho, adotamos a nomenclatura de Swales (2004).

compõem os textos desses exemplares sejam principal e majoritariamente de tipo verbal escrita, esse modo semiótico é também imagético quando interpretado como um fenômeno gráfico, ou seja, observando-se os aspectos visuais que têm funções e significados próprios (KRESS; van LEEUWEN, 2006). Esses significados estão diretamente relacionados com (mas não limitados às) intenções dos participantes da interação nos contextos de produção, circulação e consumo do gênero (FAIRCLOUGH, 1992).

A próxima subseção traz uma contextualização e descrição do TESLLE de Língua Inglesa da UFSM e os estudos que se encarregaram de reportar análises desse teste.

### **2.2.1 O TESLLE (Teste de Suficiência em Leitura em Língua Estrangeira)**

O TESLLE é aplicado semestralmente na UFSM para a comunidade acadêmica que precisa comprovar a competência em leitura acadêmica de língua adicional. Seu público-alvo são alunos de pós-graduação, para os quais a comprovação de suficiência é obrigatória, alunos de graduação que sejam prováveis formandos e que estejam no último semestre do curso e servidores (técnico-administrativos e docentes) da UFSM. Além disso, é estipulado que os examinandos falantes de português podem se inscrever no teste de língua alemã, espanhola, francesa ou inglesa, enquanto os examinandos não falantes de português podem se inscrever no teste de português como língua estrangeira (MARCUSO, 2016, p. 2). Na UFSM, o TESLLE é o principal meio de se obter a comprovação de suficiência em leitura em língua estrangeira exigida pelos programas de pós-graduação, e cerca de 1200 examinandos realizam o TESLLE por semestre (Ibid, p. 2). Essa comprovação é exigida seguindo a resolução n. 003/10 (UFSM, 2010), que estabelece as normas para realização do TESLLE na UFSM. Ainda de acordo com essa resolução, é possível realizar aproveitamentos de testes de proficiência ou suficiência de outras instituições e as revalidações, que são competência dos colegiados dos programas de pós-graduação da UFSM, que pode recorrer, se necessário, à consultoria do Departamento de Letras Estrangeiras e Modernas. A resolução também determina que serão aceitos testes que são oficialmente reconhecidos pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), podendo estes ser emitidos por

instituições com programas de pós-graduação que também sejam reconhecidas pela CAPES, e realizados no máximo há cinco anos da data do pedido de aproveitamento, ou que forem emitidos diretamente pelo MEC. Nesse sentido, o TOEFL ITP é um dos testes que tem sido utilizado como comprovação da proficiência na UFSM.

A literatura da área mostra que ao menos cinco estudos investigaram o TESLLE da UFSM. O trabalho pioneiro de Wielewicki (1997) investigou a testagem realizada em abril de 1996, e buscou, principalmente, entender a que se deviam os índices de aprovação e reprovação dos examinandos (WIELEWICKI, 1997, p. 9). Wielewicki constatou uma discrepância entre a percepção de 61 examinandos sobre seus níveis de leitura em ESP com seus conceitos (pontuação) obtidos no TESLLE. O estudo aponta que, embora a maioria (78,7%) dos examinandos tenha se autoconceituado como “bom/razoável” em leitura em ESP, o índice de reprovação destes no teste foi de 45,9%. Segundo o autor, o autoconceito dos examinandos de leitura em ESP “parece estar associado ao fato desses examinandos já estarem em estágios relativamente avançados de seus cursos, [...] ou seja, podem ser considerados leitores funcionais em ESP” (Ibid., p. 159). Enquanto o acentuado índice de reprovação foi atribuído pelo autor “à dificuldade dos examinandos em localizar informação específica no texto e em responder a questões que demandem um conhecimento mais apurado de sintaxe” (Ibid., p. 147). O estudo conclui relacionando muitas das dificuldades apresentadas pelos examinandos com as diferenças entre os conhecimentos sistêmico e esquemático necessários à leitura do texto-base do teste com os textos diretamente relacionados às suas respectivas áreas (Ibid., p. 161). O autor também levanta a possibilidade de os examinandos “não reconhecerem no teste uma medida representativa da sua habilidade de leitura, gerando problemas no que se refere à validade aparente do teste” (Ibid., p. 136).

Já a análise de Becker (2013) acerca do TESLLE concentrou-se na análise de duas edições do TESLLE que datam de 2009 e 2011. A autora identificou 10 tipos de questão: acordo/desacordo; referência; conectores lógicos; organização retórica (IMRD); extração de informação do texto; resumo; identificação da formação textual; fato/opinião; transcrição de definição; e modalidade (BECKER, 2013, p. 12). O estudo apontou que as questões de referência, de conectores lógicos e de modalidade são as em que os examinandos tiveram os maiores índices de erros.

O trabalho de Cunha (2015), por sua vez, analisou as competências exploradas pelas questões em relação às estruturas e aos tipos/enfoques<sup>7</sup> das questões. Foram analisadas duas edições do TESLLE, referentes ao segundo semestre letivo de 2013 e ao primeiro semestre de 2014, nas quais foram identificados 11 tipos de questão, a saber: relações oracionais; gênero; identificação de informação no texto; inferência; rotulação; metáfora; modalidade; multimodalidade; referência; resumo; e tradução (CUNHA, 2015, p. 12-13). Com relação às competências exploradas pelas questões, o estudo de Cunha (2015) fundamenta-se nas definições propostas por Celce-Murcia et al. (1995) e Littlewood (2004), e conclui que as questões dos TESLLE analisados estão relacionadas com as competências linguística, discursiva e pragmática.

Por sua vez, o trabalho de Radünz (2017) revisou os estudos prévios de Becker (2013) e Cunha (2015). Esse trabalho aponta que, devido à distância temporal entre as edições do TESLLE analisadas nesses estudos, “há cinco tipos de questão identificados somente no estudo de Becker, e cinco tipos de questão identificados em ambos estudos, dos quais dois são nomeados diferentemente pelas autoras” (RADÜNZ, 2017, p. 91). Essa divergência nos resultados, de acordo com Radünz (2017, p. 92), se deu pelo fato de que as edições do TESLLE analisadas por Becker (2013) e Cunha (2015) datam de um período em que o teste adotava questões discursivas e objetivas em diferentes edições, como explicado pela breve contextualização histórica do TESLLE feita pela autora:

Em 1996 o teste era composto de 15 questões relacionadas à leitura de um texto, e era dividido em cinco seções (WIELEWICKI, 1997, p. 105-106). Depois disso, o teste adotou os formatos discursivo e objetivo em diferentes edições, até 2014, quando foi padronizado para o formato múltipla-escolha novamente, de acordo com informações fornecidas pela comissão de organização do TESLLE (RADÜNZ, 2017, p. 90-91).

Mais recentemente, outros dois trabalhos relataram estudos acerca do TESLLE como um gênero discursivo. O já mencionado artigo de Radünz e Marcuzzo (2017) buscou relatar a relação entre os resultados dos examinandos com os tipos e os enfoques das questões de seis diferentes edições do TESLLE de Língua Inglesa (Ibid., p. 89). Em

---

<sup>7</sup> Os termos “enfoques” e “tipos” de questão são empregados como sinônimos no trabalho de Cunha (2015).

seguida, outro artigo reportou uma Análise Crítica de Gênero das mesmas edições do TESLLE (MARCUIZZO; RADÜNZ, 2019). Em ambos os trabalhos, as autoras diferenciam tipos e enfoques de questão. Com base em Medeiros (1977), as questões foram classificadas em cinco tipos: completamento; certo-ou-errado; escolha múltipla; associação; e ordenação. Os enfoques linguísticos das questões, por sua vez, foram classificados em 12 categorias: advérbios, evidência textual, figura de linguagem, funções de partes do texto, gênero, inferência, modalidade, multimodalidade, referência, relações oracionais, tempos e vozes verbais e tradução (MARCUIZZO; RADÜNZ, 2019).

Na próxima subseção buscamos contextualizar e descrever o TOEFL ITP e os estudos recentes que se concentraram nesse teste, em especial os que enfatizaram análises da seção de leitura desse teste.

### **2.2.2 O TOEFL ITP (*Test of English as a Foreign Language - Institutional Testing Program*)**

O TOEFL é uma marca registrada da ETS (*Educational Testing Service*), que é uma organização sem fins lucrativos que desenvolve avaliações e testes educacionais. A ETS atua principalmente nos Estados Unidos, mas também administra testes internacionais em mais de 180 países, como o TOEFL (*Test of English as a Foreign Language*), o TOEIC (*Test of English for International Communication*), e o GRE (*Graduate Record Examination*) (ETS, 2020a).

As diferentes modalidades dos testes TOEFL são utilizadas como instrumentos de avaliação em diversos contextos/países e para uma vasta gama de finalidades. Em muitos países e em instituições o TOEFL é utilizado por sua reputação de confiabilidade. No Brasil, os testes TOEFL são ofertados principalmente em duas modalidades: ITP (*Institutional Testing Program*) e iBT (*Internet Based Test*). A versão ITP é a mais comumente escolhida internacionalmente por seu custo mais acessível em comparação com a versão da modalidade iBT (TAUFIQ; SANTOSO; FEDIYANTO, 2018, p. 227), bem como por sua maior praticidade de aplicação em relação a outros testes de proficiência.

O TOEFL ITP é dividido em dois níveis: i) *Level 1* - básico a intermediário; e ii) *Level 2* - intermediário a avançado. A versão ITP (*Level 2*), que foi analisada no presente

estudo, é a modalidade impressa em papel que aferi proficiência em inglês como língua estrangeira, enquadrando a pontuação do teste (de 310 a 677 pontos) entre os níveis A2 e C1 do Quadro Comum Europeu de Referência para Idiomas<sup>8</sup> (QCER, ou apenas QCE). O QCER é um padrão internacionalmente reconhecido de descrição de proficiência em idiomas. O QCER estabelece seis níveis de proficiência, que são, respectivamente, A1, A2, B1, B2, C1 e C2, em que A1 indica nível mais básico e C2 o mais proficiente em determinada língua.

Os resultados dos testes TOEFL são utilizados para diversas finalidades, dentre elas, para aferir e comprovar proficiência em inglês para fins pessoais e/ou profissionais, especialmente por acadêmicos com interesse em realizar mobilidade acadêmica internacional ou cumprir requisitos de editais de diversos tipos de seleções.

No Brasil, a versão ITP do TOEFL é distribuída exclusivamente pela Mastertest, empresa que também é responsável pelo credenciamento e gerenciamento de todos os centros aplicadores em território nacional, que incluem cursos livres de línguas, universidades públicas e privadas, órgãos governamentais, e empresas privadas que utilizam esses testes em programas na área de recursos humanos. A empresa ainda gerencia a aplicação desse teste em mais de 150 universidades públicas federais e estaduais, onde mais de meio milhão de testes já foram aplicados, segundo informações do *site* da Mastertest<sup>9</sup>.

Além disso, o TOEFL ITP se tornou consideravelmente popular no Brasil quando o programa IsF passou a ofertar esse teste (bem como cursos voltados para o teste) gratuitamente nas universidades e instituições públicas em que o programa foi implementado. O programa IsF, ao todo, aplicou mais de 450 mil testes TOEFL ITP durante sua existência, entre 2012 e 2017. Além disso, o programa IsF ofertou, ao todo, mais de 350 mil vagas em cursos de idiomas, das quais mais de 330 mil foram para cursos de língua inglesa<sup>10</sup> que frequentemente tinham esse teste como pré-requisito de nivelamento e/ou preparavam alunos para realizar o teste TOEFL ITP. O grande impacto do IsF foi responsável por tornar o teste TOEFL ITP amplamente conhecido pela

---

<sup>8</sup> Em inglês: *Common European Framework of Reference for Languages (CEFR)*.

<sup>9</sup> Endereço: <https://mastertest.com.br/>. Acesso em: 15 fev. 2020.

<sup>10</sup> Fonte: <http://isf.mec.gov.br/historico-botoes/pesquisas-e-relatorios/33-pesquisas-e-relatorios/199-relatorio-nacional>. Acesso em 20 de maio de 2021.

comunidade acadêmica brasileira para seus diferentes usos. No contexto da UFSM, por exemplo, o TOEFL ITP passou a ser comumente adotado como instrumento para a comprovação de proficiência em língua estrangeira pelos alunos de pós-graduação da universidade em função disso.

Também por conta dessa popularidade internacional, os testes TOEFL são objeto de numerosos estudos em vários países. O presente levantamento da literatura<sup>11</sup> indica que, nos últimos anos, o TOEFL ITP foi objeto de pesquisa de, pelo menos, seis estudos nos pouco mais de 10 anos atrás: Kamijo (2009), Perez-Amurao (2011), Santos (2016), Silahahi (2017), Marcuzzo e Azambuja (2017) e Taufiq, Santoso e Fediyanto (2018).

Kamijo (2009) analisou uma seção de leitura do TOEFL ITP a partir de um livro didático preparatório, de forma a descrever sua estrutura para propósitos pedagógicos. O estudo demonstrou que essa seção é definida em conformidade com o contexto do inglês para fins acadêmicos (*English for Academic Purposes - EAP*)<sup>12</sup> e que seus textos-base são de tipo histórico/narrativo (KAMIJO, 2009, p. 204), além de identificar múltiplos enfoques<sup>13</sup> de questão: busca por informações específicas (*scanning*), compreensão de vocabulário em contexto, compreensão de ideias principais e detalhes de apoio, e inferência de afirmações do texto (Ibid., p. 198).

Perez-Amurao (2011) identificou o tipo (enfoque) e a natureza (grau de dificuldade) das questões da seção de leitura do TOEFL ITP (a que chamou de sub-teste) (PEREZ-AMURAO, 2011, p. 44). A partir da análise de um material preparatório, o estudo classificou as questões da seção de leitura em sete tipos, e as classificou em quatro níveis (extremamente difícil, difícil, moderadamente difícil e fácil) de acordo com o percentual de acerto de 43 alunos que participaram da pesquisa (Ibid., p. 44). Os tipos de questões identificados foram, em ordem decrescente de dificuldade: vocabulário, fatos e detalhes, tom e propósito, inferência, ideia principal, organização e referentes (Ibid., p. 53).

---

<sup>11</sup> Para o levantamento da literatura deste trabalho, realizaram-se pesquisas por publicações científicas a partir da busca pelos termos TOEFL ITP, *Reading e Critical Genre Analysis* na ferramenta *Google Scholar* no período de fevereiro de 2020. Foram selecionados estudos do TOEFL ITP e seu respectivo teste de leitura que foram publicados nos últimos 10 anos.

<sup>12</sup> Neste trabalho, entende-se que o EAP é uma vertente do ESP.

<sup>13</sup> Kamijo (2009) utiliza o termo "formatos de questão". Entretanto, neste trabalho, entendemos esse conceito como enfoques, uma vez que este não diz respeito à estrutura, mas aos conteúdos mobilizados pelas questões.

Santos (2016) avaliou o que chamou de “compreensão leitora” no teste TOEFL ITP e propôs reflexões sobre possíveis efeitos retroativos a partir do uso das estratégias de leitura por candidatos na seção de leitura do TOEFL ITP. A partir da aplicação de questionários e realização de entrevistas, o estudo mostrou que “o conhecimento e [a] aplicação das estratégias de leitura[,] podem contribuir com um bom desempenho do candidato na seção de leitura do teste” (SANTOS, 2016, p. 9). Além disso, como um dos efeitos retroativos (*washback effect*) do teste, “os candidatos que não tiveram um bom rendimento mostraram-se determinados a aprimorar o desenvolvimento da habilidade de leitura” e a estudar estratégias que possam ajudar em uma realização futura do teste. (SANTOS, 2016, p. 9).

Silahahi (2017) avaliou a habilidade de pensamento crítico de 140 alunos de universidade por meio de suas pontuações em um modelo do teste de leitura do TOEFL ITP. Ao explorar as características de pensamento crítico e de competência de leitura, é demonstrado que ambas são inseparavelmente relacionadas (SILAHAHI, 2017, p. 81). O estudo conclui que, para que se obtenha um melhor desempenho em testes de leitura como o TOEFL ITP, faz-se necessário integrar as habilidades de leitura às de pensamento crítico, e que isso pode ser alcançado com a prática do teste e com o auxílio de professores de língua inglesa.

Marcuzzo e Azambuja (2017) analisaram todas as três seções do TOEFL ITP a partir do *Official TOEFL Handbook* de 2012 (MARCUIZZO; AZAMBUJA, 2017). Nesse estudo, a seção de leitura foi analisada de forma a identificar os gêneros dos textos-base, as competências exploradas pelas questões e a perspectiva de leitura da seção (MARCUIZZO; AZAMBUJA, 2017, p. 346). As autoras constataram que os textos-base dessa seção não incluem fonte ou autoria, além de não manterem o *layout* de suas publicações originais, dessa forma, os textos-base parecem estar em desacordo com o que se entende por texto como exemplar de gênero (SWALES, 1990, p. 347). O estudo ainda identificou que as competências exploradas nessa seção são de tipo linguística e pragmática (Ibid., p. 347), e que as questões que exploram aspectos sociolinguísticos e socioculturais não são exploradas no teste, que também parece não abordar, nas questões, aspectos relacionados à perspectiva de gênero (Ibid., p. 349).

Mais recentemente, Taufiq, Santoso e Fediyanto (2018) conduziram uma pesquisa qualitativa que analisou criticamente o TOEFL ITP como um instrumento de avaliação (TAUFIQ; SANTOSO; FEDIYANTO, 2018, p. 226). O estudo avaliou a “praticidade” do teste, ou seja, a viabilidade de se “conduzir o teste sem problemas processuais” (Ibid., p. 227, tradução do autor).

Foi constatado que o teste em si é suficientemente prático, uma vez que “os detalhes de administração [do teste] são previamente estabelecidos [...] bem como os materiais e equipamentos” (Ibid., p. 227, tradução nossa). No entanto, os autores apontam que devem-se observar alguns aspectos da aplicação, como o número de examinandos que realizam o teste simultaneamente, e garantir a equidade de condições e recursos disponíveis a estes, como o acesso a salas e equipamentos (Ibid., p. 228).

Além disso, em conformidade com as constatações de Marcuzzo e Azambuja (2017), algumas habilidades, como as de produção oral e escrita, não são testadas pela modalidade ITP, fazendo com que a proficiência testada pelo TOEFL ITP seja apenas em inglês “passivo” (TAUFIQ; SANTOSO; FEDIYANTO, 2018, p. 227). Isso implica que examinandos que tenham uma melhor proficiência nessas habilidades podem ser prejudicados por não terem a oportunidade de demonstrá-las no teste, ou levar à necessidade de outros testes serem aplicados (Ibid., p. 227-228). Por fim, a autenticidade do TOEFL ITP, também analisada pelo estudo, só é questionada pelo fato de os itens desse teste serem previamente testados em outros testes TOEFL. Esse fato, segundo os autores, torna as modalidades iBT e PBT mais “genuínas” (Ibid., p. 228), no sentido em que os itens desses testes não são utilizados em outros testes TOEFL, como acontece com o TOEFL ITP.

Cabe também citar o estudo de Samad, Jannah e Fitriani (2017), que, similarmente ao presente estudo, tem como *corpus* o teste de leitura do TOEFL, ainda que na modalidade PBT. Os autores realizaram uma análise quantitativa para investigar as dificuldades e as estratégias utilizadas por alunos de graduação ao realizar uma amostra desse teste (SAMAD; JANNAH; FITRIANI, 2017, p. 31). As respostas incorretas de 30 alunos que realizaram o teste foram classificadas em 13 habilidades, que correspondem a: responder questões implícitas (26%), responder questões de detalhes declarados corretamente (14%), usar o contexto para compreender o significado de palavras difíceis

(12%), responder questões de ideia principal corretamente (11%), descobrir o significado a partir de partes de palavras (9%), encontrar detalhes não-declarados (8%), reconhecer a organização das ideias (5%), usar o contexto para compreender o significado de palavras simples (4%), encontrar referentes pronominais (3%), procurar por informações específicas (3%), responder questões de transição (2%), descobrir o significado de pistas estruturais (2%), e identificar tom, propósito ou curso (1%). Ao relacionar esses dados com as respostas dos questionários aplicados aos alunos, a investigação aponta que as estratégias utilizadas por eles são bastante variadas, e que aliá-las a tipos específicos de habilidades de leitura mobilizados pelas questões é de fundamental importância para um desempenho satisfatório no teste.

Apesar da relevância dos estudos apresentados nesta seção, não encontramos estudos que analisem a seção de leitura do TOEFL ITP como um gênero discursivo inserido em uma prática social, muito embora esse tipo de análise já tenha sido feito, em certa medida, por Perez-Amurao (2011), Marcuzzo e Azambuja (2017) e Taufiq, Santoso e Fediyanto (2018). A interpretação da seção de leitura (bem como das demais seções) do TOEFL ITP como gênero discursivo parece ser pertinente na medida em que há uma tipificação estrutural reconhecível em termos de pertinência linguística, eventos e participantes envolvidos na atividade comunicativa específica dessa seção. Além disso, o gênero que a seção de leitura constitui é discursivo na medida em que ela compreende uma atividade humana (a testagem) constituída pela linguagem (de língua estrangeira) em um contexto social específico (âmbito dos testes de EAP).

Na seção 2.4 são descritas e discutidas as categorias de análise das questões dos testes do *corpus* deste estudo.

### 2.3 ENFOQUES E TIPOS DE QUESTÃO

A análise dos enfoques das questões está fundamentada no estudo de Marcuzzo e Radünz (2019). O termo enfoque diz respeito aos aspectos da linguagem que, de alguma forma e/ou em alguma medida, são direta ou indiretamente mobilizados pelas questões. Ao analisar seis edições do TESLLE, Marcuzzo e Radünz (2019) identificaram 12 enfoques de questão.

As edições do TESLLE de 2014 a 2016 enfocam: **1) Advérbios** – é esperado que o examinando identifique diferentes advérbios de tempo, modo e espaço; **2) Evidência textual** – o mesmo que “identificação de informação no texto” por Cunha (2015, p. 12) – nessas questões, é esperado que o examinando identifique evidências textuais para responder o que está sendo perguntado; **3) Figura de linguagem** – “metáfora” por Cunha (2015, p. 12) – é esperado que o examinando identifique a linguagem metafórica, ou seja, quando os autores usam uma palavra, expressão ou oração para se referir a outro significado que não o significado literal de uma palavra, expressão ou oração; **4) Funções de partes do texto** – “rotulação” por Cunha (2015, p. 12) – é esperado que o examinando rotule alguns segmentos, considerando se eles estão sugerindo, questionando ou exemplificando; **5) Gênero** – é esperado que o examinando identifique funções de diferentes partes do texto e/ou o propósito comunicativo do gênero; **6) Inferência** – é esperado que o examinando infira e avalie o significado de frases específicas; **7) Modalidade** – é esperado que o examinando identifique os diferentes níveis de modalidade em partes específicas do texto; **8) Multimodalidade** – é esperado que o examinando identifique a relação da informação nas imagens (no texto não verbal) com a informação no texto verbal; **9) Referência** – é esperado que o examinando identifique os referentes exofóricos, anafóricos e catafóricos de pronomes específicos, grupos nominais e pronomes, tais como “it”, “we”, “you”, “this”, “the last of these”, etc; **10) Relações oracionais** – é esperado que o examinando identifique o tipo de relação entre as orações; **11) Tempos e vozes verbais** – é esperado que o examinando identifique diferentes tempos e vozes verbais em inglês em orações específicas; e **12) Tradução** – é esperado que o examinando traduza um trecho do texto em inglês para o português. No corpus, duas questões envolvem dois enfoques. Uma mistura evidência textual e modalidade, já a outra, evidência textual e multimodalidade. (Ibid. p. 3648. Grifos nossos).

Além disso, Marcuzzo e Radünz (2019) também relacionaram os enfoques das questões com os estratos da linguagem (HALLIDAY, 1978; HALLIDAY; HASAN, 1989) que elas “acionam”, os quais são, do nível mais concreto para o mais abstrato, respectivamente: Fonologia e Grafologia; Léxico-gramática; Semântica e Pragmática; Registro; Gênero; e Discurso<sup>14</sup>.

Quadro 2 - Relação entre as competências, estratificação da linguagem e enfoques das questões

Competência	Estratificação	Enfoques das questões
Linguística	Léxico-gramática	Relações oracionais Tempos e vozes verbais Advérbios
Discurso	Gênero Ideologia	Gênero Modalidade

<sup>14</sup> O uso de letras maiúsculas para cada estrato da linguagem é uma convenção da Linguística Sistemico-Funcional quando estes referem-se a categorias de análise linguística.

Pragmática	Semântica e pragmática	Evidência textual Referência Inferência Figura de linguagem Funções de partes do texto Multimodalidade Tradução
------------	------------------------	---

Fonte: Adaptado de Radünz e Marcuzzo (2017, p. 107).

Paralelamente, o termo “tipo de questão” diz respeito a como as questões dos testes são estruturadas com relação a seus formatos, isto é, de que forma elas são articuladas de modo a solicitar ou demandar determinados tipos de resposta. Nesse sentido, ao analisar as questões de testes, Medeiros (1977) as classificou em cinco tipos: 1) completamento; 2) certo-ou-errado; 3) escolha múltipla; 4) associação; e 5) ordenação. Nas questões de completamento, de acordo com Medeiros (Ibid., p. 37), o examinando deve completar segmentos que estão incompletos, e há somente uma resposta possível para esse tipo de questão. Já nas questões de certo-ou-errado, o examinando deve considerar se as afirmações apresentadas estão certas ou erradas (Ibid., p. 45). As questões de escolha múltipla fornecem possíveis respostas para a pergunta apresentada, e o examinando deve identificar a correta (Ibid. p. 53). As questões de associação apresentam dois grupos de informações que devem ser associados (Ibid., p. 79). Por último, questões de ordenação são aquelas que pedem para que o examinando coloque elementos de um grupo em ordem (Ibid., p. 87). Nessa perspectiva, o estudo mais recente acerca do TESLLE identificou três tipos de questão, que são: i) de pergunta e resposta; ii) de acordo ou desacordo; e iii) de associação (MARCUIZZO; RADÜNZ, 2019, p. 3647).

Essas categorias de análise textual integram e ajudam a consolidar a Análise Crítica de Gênero, perspectiva teórico-metodológica implementada no presente estudo e que é apresentada e discutida na seção 2.4.

## 2.4 A ANÁLISE CRÍTICA DE GÊNERO

A ACG configura-se como uma orientação interdisciplinar brasileira para estudos de gênero que foi pensada e proposta inicialmente por Meurer (2002; 2003; 2004), e

posteriormente elaborada e implementada, principalmente no Brasil, por Motta-Roth (2005; 2008) e Heberle (MOTTA-ROTH & HEBERLE, 2015). A ACG surgiu da latente necessidade de desenvolver uma abordagem que contemplasse e correlacionasse investigações de gênero a partir de análises dos níveis léxico-gramatical, semântico, discursivo, contextual, sociocultural e ideológico. Esses níveis são análogos à estratificação dos planos comunicativos da linguagem (MARTIN, 1992), conforme apresenta a Figura 2, onde cada círculo concêntrico ilustra um estrato da linguagem. Esses estratos são, do mais externo para o mais interno: Discurso, Gênero, Registro, Semântica e Pragmática, Léxico-gramática, Fonologia e Grafologia, respectivamente. Dessa forma, os estratos mais externos instanciam e são conceitual e linguisticamente mais abstratos que os estratos internos, que, por sua vez, são progressivamente mais concretos e realizam os estratos mais externos. Além disso, os graus de nitidez das cores utilizadas em cada círculo concêntrico do diagrama ilustram o grau de abstração (tons mais claros) e concretude (tons mais escuros) de cada estrato. Por sua vez, a ausência de contraste nas linhas que marcam cada estrato demonstra a indissociabilidade prática entre texto, interação e contexto, que são conceitualizados para fins teórico-metodológicos.

Figura 2 - Articulação das concepções de discurso como texto, interação e contexto<sup>15</sup> e de linguagem como sistema sociosemiótico<sup>16</sup>

---

<sup>15</sup> (FAIRCLOUGH, 1989, p. 25).

<sup>16</sup> (MARTIN, 1992, p. 496; MOTTA-ROTH, 2008, p. 355).



Fonte: Motta-Roth e Scherer (2016, p. 169).

A estratificação da linguagem nos fornece clareza e embasamento teórico para identificar que aspectos da linguagem estão sendo mobilizados em algum grau por determinadas amostras de linguagem. No presente estudo, esses estratos embasaram a análise dos enfoques das questões.

O estrato do Discurso diz respeito “a(s) ideologia(s), ou seja, as visões de como as ‘coisas’ deveriam funcionar” (NASCIMENTO, 2017, p. 129). O estrato do Gênero concerne os “usos da linguagem em práticas discursivas relativamente estáveis e reconhecidas em uma cultura” (Ibid., p. 230). O estrato do Registro trata da forma como aspectos de um contexto de situação são refletidos em registros, ou seja, instanciações específicas de gêneros discursivos, revelando especificidades do contexto das relações sociais envolvidas no texto e sua organização estrutural (Ibid., p. 230). O estrato da Semântica e Pragmática se refere aos significados e à sistematização dos elementos linguísticos, de forma a construir e representar linguisticamente aspectos da experiência humana no texto (metafunção ideacional da linguagem), estabelecer relações entre interlocutores (metafunção interpessoal), e atribuir valores às informações de acordo com suas distribuições no texto (metafunção textual) (Ibid., p. 231). Os estratos da Léxico-

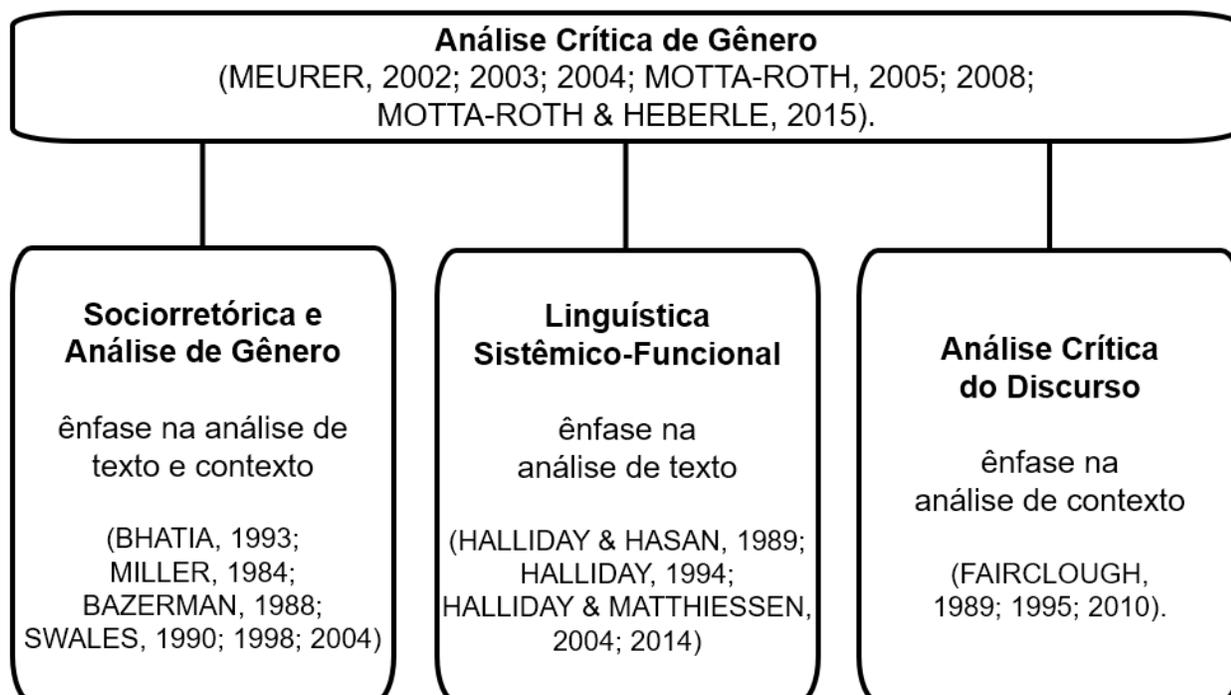
gramática e da Fonologia e Grafologia dizem respeito às estruturas linguísticas utilizadas para construir esses significados (na forma de frases, complexos oracionais, orações, grupos nominais, palavras, sílabas, grafemas, fonemas, etc.).

Essa concepção de estratificação da linguagem é fundamental para a ACG, abordagem que “leva em consideração a léxico-gramática, o discurso e o contexto sociocultural, relacionando forma e função, discurso e ideologia, a partir de uma perspectiva analítica do discurso crítico” (MOTTA-ROTH; HEBERLE, 2015, p. 2).

Dessa forma, estudos de ACG partem do pressuposto de que a relação entre as práticas sociais e os textos que as constituem e realizam é bidirecional, ou seja, textos e práticas sociais são constituídos e constituem um ao outro. Assim sendo, para que se faça uma ACG, pode-se usar tanto o(s) texto(s) (gêneros discursivos) quanto seu(s) contexto(s) como ponto de partida, oscilando entre um e outro conforme conveniente ou necessário, de forma a entender as relações indissociáveis entre estes. Assim sendo, estudos que têm como embasamento teórico-metodológico a ACG têm objetivos crítico-analíticos, ou seja, empenham-se em identificar, descrever e explicar como a linguagem, em todos os seus contextos e instâncias, materializa, articula e dissemina discursos enquanto posições ideológicas (HENDGES, 2019, p. 2).

Conforme ilustra a Figura 3, a ACG congrega contribuições interdisciplinares da Análise de Gênero, especialmente aquelas relacionadas ao ESP (BHATIA, 1993; SWALES, 1990; 1998; 2004), Sociorretórica (MILLER, 1984; BAZERMAN, 1988), Linguística Sistêmico-Funcional (LSF) (HALLIDAY; HASAN, 1989; HALLIDAY, 1994; HALLIDAY; MATTHIESSEN, 2004; 2014) e Análise Crítica do Discurso (ACD) (FAIRCLOUGH, 1989; 1995; 2010).

Figura 3 - Aportes teórico-metodológicos da Análise Crítica de Gênero



Fonte: elaborado pelo autor com base em Motta-Roth e Heberle (2015).

Da Sociorretórica e da Análise de gênero, a ACG “se apropria da visão de como o gênero funciona dentro de instituições sociais e envolve disparidades de poder, ideologia e identidade” (MOTTA-ROTH; HEBERLE, 2015, p. 5).

A ACG também utiliza a LSF para a análise da léxico-gramática a fim de identificar as representações de mundo e experiência (metafunção ideacional), as manifestações semióticas das relações sociais (metafunção interpessoal), e os significados oriundos da organização dessas representações em um texto (metafunção textual).

Da ACD, a ACG incorpora o modelo tridimensional proposto por Fairclough (2001) que inclui três níveis de análise:

- 1) análise de escolha lexical, gramática, coesão e estrutura dos textos, para descrever um texto enquanto um evento discursivo;
- 2) análise de interdiscursividade, intertextualidade, coerência e força nos contextos de produção, distribuição e consumo de textos para a interpretação das práticas discursivas;
- 3) análise de ideologia e hegemonia para a explicação do contexto das práticas sociais (MEURER, 2005, p. 95).

Assim, a LSF possibilita investigar os estratos da fonologia, grafologia, semântica e pragmática do contexto de situação, enquanto a análise de gênero de Swales e a Sociorretórica viabilizam recursos para averiguação do contexto de cultura, e, por sua vez, a ACD possibilita investigar o estrato do discurso (FLOREK, 2018, p. 69).

Para que possamos entender melhor os princípios da ACG, nos voltamos às concepções de linguagem, leitura e letramento que fundamentam essa abordagem e seus aportes teórico-metodológicos.

## 2.5 LINGUAGEM, LEITURA E (MULTI)LETRAMENTO(S)

Para que possamos dar conta de identificar as perspectivas de linguagem, leitura e letramento subjacentes aos testes e às práticas sociais aqui investigadas, precisamos, primeiramente, buscar entender essas concepções.

Neste estudo, a linguagem é entendida conforme os pressupostos nos quais a ACG se embasa. Dentro dessa perspectiva, a linguagem é definida como um “um fato social” (HALLIDAY, 1978, p. 1) e um “recurso de produção de significado” (HALLIDAY; MATTHIESSEN, 2014, p. 3). Uma vez que a linguagem e o contexto são vistos como interdependentes (THOMPSON, 2004, p. 9), a linguagem necessariamente envolve uma dimensão sociocultural. Nesse sentido, a linguagem pode ser

investigável em termos de: i) função, aquilo que fazemos ou podemos fazer com ela; ii) semântica, o sentido que produzimos a partir dela; iii) contexto, como os sentidos são influenciados pelas situações sociais e culturais em que são produzidos; e iv) semiótica, como produzimos sentidos pela seleção de recursos potencialmente significativos (som, cor, imagem, linguagem verbal escrita/oral, p. ex.) que integram uma gama de modos de significação (HALLIDAY, 1978; HALLIDAY; HASAN, 1989; apud. FLOREK, 2018, p. 62).

Leitura, por sua vez, é uma habilidade linguística cujo propósito principal é “obter significado de um texto” (NUTTALL, 1996, p. 4). Nesse sentido, a autora destaca que a ideia de leitura não é uma atividade meramente passiva por parte do leitor (Ibid., p. 5). Ao contrário, trata-se de um processo interativo entre autor e leitor, no qual ambos dependem um do outro para o sucesso dessa interação (Ibid., p. 11).

Por fim, letramento, que é a tradução para o português brasileiro feita por Soares (1998) do termo *literacy*, que foi inicialmente proposto por Street (1984), diferencia-se das demais possíveis traduções do termo, como, por exemplo, alfabetização. Alfabetização “tem um foco individual, bastante ditado pelas capacidades e competências (cognitivas e linguísticas) escolares e valorizadas de leitura e escrita” (ROJO, 2009, p. 98), enquanto letramento é definido como “o resultado da ação de ensinar ou de aprender a ler e escrever: estado ou a condição que adquire um grupo social ou indivíduo como consequência de ter se apropriado da escrita” (SOARES, 1998, p.18) e “como um conjunto de práticas sociais que usam a escrita, enquanto sistema simbólico e enquanto tecnologia, em contextos específicos, para objetivos específicos” (KLEIMAN, 1995, p. 19).

Marques (2015, p. 31) aponta que essas concepções de letramento se assemelham à concepção de alfabetização proposta por Freire (1979):

é mais que o simples domínio psicológico e mecânico de técnicas de escrever e de ler. É o domínio destas técnicas em termos conscientes. [...] Implica numa autoformação de que possa resultar uma postura interferente do homem sobre seu contexto (FREIRE, 1979, p. 110).

Em virtude das incontáveis tecnologias que surgiram e/ou se popularizaram em todo o mundo por volta dos anos 90, o termo letramentos foi “atualizado” para multiletramentos. A palavra multiletramentos incorpora o conjunto múltiplo de habilidades que se faz necessário para que se possa compreender e participar ativamente da vasta gama de gêneros multimodais dos quais fazemos parte enquanto indivíduos e sociedade. Esse termo busca destacar não somente a importância, mas também a necessidade de que sejam pensadas e implementadas pedagogias de “letramentos multissemióticos exigidos pelos textos contemporâneos, [que] ampliam a noção de letramentos para o campo da imagem, da música, das outras semioses que não somente a escrita” (ROJO, 2009, p. 107), como também defende o *New London Group*:

Decidimos que os resultados de nossas discussões poderiam ser encapsulados em uma palavra - multiletramentos - uma palavra que escolhemos para descrever dois importantes argumentos que podemos ter com a ordem cultural emergente, institucional e global: a

multiplicidade de canais de comunicação de mídias, e a proeminência crescente da diversidade cultural e linguística (New London Group, 1996, p. 63).

A partir desses conceitos, fundamentais para pesquisas de linguagem, e ao relacionar as análises de texto e contexto, torna-se possível delinear com precisão as concepções de linguagem, proficiência em leitura de EAP e (multi)letramento(s) que são subjacentes à prática social aqui estudada. Essas concepções são importantes pois, ainda que sejam naturalizadas e relativamente intrínsecas aos testes e à prática social aqui investigados, exercem influência sobre outras práticas sociais, como as de ensino, por exemplo. Essa discussão é retomada em mais detalhe nos capítulos finais deste estudo.

Na próxima seção apresentamos os conceitos de confiabilidade, validade e efeito retroativo, que consideramos fundamentais para entender a coerência, a importância e os papéis exercidos pelos testes e(m) suas práticas de testagem.

## 2.6 CONFIABILIDADE, VALIDADE E EFEITO RETROATIVO

Os conceitos de confiabilidade (*reliability*, também traduzido para fidedignidade) e validade (*validity*) de testes são estreitamente relacionados um com o outro. De maneira geral, pode-se dizer que a confiabilidade de um teste “está diretamente relacionada com seu grau de precisão como instrumento de mensura[ç]ão de uma determinada habilidade linguística” (CÉLIA, 1983, p. 42 apud WIELEWICKI, 1997, p. 60), enquanto que o conceito de 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; CLAPHAM; WALL, 1995, p. 89)<sup>17</sup>

---

<sup>17</sup> Em inglês: “*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*”. Tradução de Scaramucci, 2009.

Wielewicki (1997) cita alguns fatores que podem afetar a confiabilidade de um teste:

acredita-se que basicamente quatro grupos de fatores afetem o escore de um candidato num teste: habilidade comunicativa de língua, características do método de testagem, atributos pessoais (por exemplo, conhecimento prévio na área de conteúdo testada) e fatores aleatórios (como, por exemplo, o nível de tensão do examinando) WIELEWICKI, 1997, p. 61).

De todo modo, o autor também afirma que “[...] é necessário observar que, em função dos diferentes fatores que podem influenciar o desempenho dos examinandos num teste, é virtualmente impossível criar um teste perfeitamente fidedigno [confiável]” (ALDERSON, CLAPHAM; WALL, 1995, p. 87, apud WIELEWICKI, 1997, p. 63).

Scaramucci (2009) faz uma brilhante contextualização histórica e discussão acerca do complexo conceito de validade. A autora explica que a concepção tradicional validade passou por uma reformulação a partir do final das décadas de 1980 e ao longo da década de 1990, principalmente em virtude do impacto da publicação intitulada *Validity*, de Messick (1989). O conceito de validade passou então a ser expandido para incorporar as dimensões e consequências sociais dos processos de testagem (SCARAMUCCI, 2009, p. 32-33).

A crítica principal ao conceito tradicional de validade era o fato de ser fragmentado e incompleto, elaborado exclusivamente por especialistas em medidas, seguindo uma visão essencialmente psicométrica e, como tal, não levar em conta, como base para a ação, as implicações de valor do significado dos resultados ou escores, nem as consequências sociais do uso desses resultados, ou seja, a dimensão social e política que devia estar presente na avaliação de línguas pelo fato de ser uma prática social. O novo conceito, embora contemplando múltiplas facetas, unifica-se em torno da validade de construto, passando a considerar não apenas bases evidenciais, mas também as consequenciais (SCARAMUCCI, 2009, p. 32).

Chapelle (1999) explica essa mudança no conceito de validade relacionando-o com o conceito de confiabilidade. A confiabilidade de um teste, que era vista como distinta e uma condição necessária para a validade, passa a ser vista como um tipo de evidência de validade. Dessa forma, validade, que era considerada meramente como a medida em

que um teste mede aquilo que pretende medir, passa a ser entendida como “um argumento relativo à interpretação e [ao] uso: o grau em que as interpretações e usos de um teste podem ser justificados” (CHAPELLE, 1999, p. 258).

Também relacionando as concepções de validade e de confiabilidade, Scaramucci (2009) explica que, a fim de ser válido, um teste “precisa avaliar com precisão e de forma consistente” (SCARAMUCCI, 2009, p. 32), ou seja, precisa ser confiável.

Se uma prova [teste] é corrigida por dois corretores e as notas obtidas são completamente distintas (10 e 0), por exemplo, que resultado devemos tomar como evidência daquilo que pretendemos avaliar? Um teste confiável pode não ser válido, entretanto. [...] Um teste de desempenho de produção escrita que, por outro lado, tem por objetivo a redação de um texto pode ser válido, embora não necessariamente confiável, se não forem estabelecidos critérios claros para a correção, se os corretores não forem treinados para a tarefa, e assim por diante. Dessa forma, um aumento de validade geralmente leva a uma diminuição de confiabilidade e vice-versa, revelando a tensão existente entre os dois parâmetros (Ibid., p. 32).

Em testes de leitura, essa “tensão” entre os conceitos de confiabilidade e validade se dá de forma similar. Ao contrário de um teste de produção escrita, como o exemplo de Scaramucci (2009), a compreensão escrita (leitura) não resulta necessariamente em uma produção textual que pode ser avaliada de maneira objetiva. Por consequência, a leitura de um examinando em um teste de proficiência é observada a partir da noção de construto, que

[...] 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 (EBEL; FRISBIE, 1991, p. 108)<sup>18</sup>.

No entanto, questões objetivas de múltipla-escolha, como as do TESLLE e do TOEFL ITP, se utilizadas de maneira pertinente, podem contribuir para que esses testes sejam suficientemente confiáveis e válidos para que cumpram os propósitos sociais a que se propõem.

---

<sup>18</sup> “The term construct refers to a psychological construct, a theoretical conceptualisation about an aspect of human behaviour that cannot be measured or observed directly. Examples of constructs are intelligence, achievement motivation, anxiety, achievement, attitude, dominance, and reading comprehension.” Tradução de Scaramucci, 2009.

Com vistas a discutir a possibilidade de que se estabeleça um equilíbrio entre confiabilidade e validade, Wielewicki (1997) aponta uma série de sugestões feitas por Hughes (1989) com a finalidade de aprimorar a confiabilidade de testes.

Entre elas, pode-se citar o uso de uma amostragem de comportamentos o mais ampla possível, o exercício de um controle rígido sobre as tarefas propostas, evitando assim redundância e ambiguidade nas instruções e nos itens incluídos no teste, a familiarização dos candidatos com o procedimento e as técnicas de testagem, a pontuação de forma objetiva feita por pessoal especificamente treinado para tal fim e a discussão prévia sobre quais respostas são aceitáveis e quais escores são apropriados. O autor adverte, no entanto, que ao buscar maior fidedignidade [confiabilidade] nos testes desenvolvidos, devemos estar cientes do risco da redução de sua validade (Ibid., p. 42) (WIELEWICKI, 1997, p. 62-63).

A confiabilidade e a validade também são interpretadas como fatores que contribuem para o efeito retroativo (em inglês, *backwash* ou *washback effect(s)* ou simplesmente *backwash* ou *washback*)<sup>19</sup> de testes e suas práticas de testagem. O conceito de efeito retroativo é frequentemente definido em relação as influências potenciais que testes exercem em seus contextos educacionais, como a definição de Scaramucci (2004, p. 203): “[o]bservações sobre o impacto ou influência que exames ou testes e avaliação em geral exerce potencialmente nos processos educacionais, seus participantes e produtos do ensino e aprendizagem”. No entanto, a própria autora explica que o efeito retroativo de testes vai além do âmbito educacional, e defende que esse conceito deve “ser avaliado com referência às variáveis, [aos] objetivos e [aos] valores da sociedade, do sistema educacional em que é usado, assim como dos resultados potenciais de seu uso” (SCARAMUCCI, 2004, p. 204-205).

Scaramucci também explora a complexidade do conceito de efeito retroativo e lista as várias dimensões deste na literatura:

a) *especificidade* (geral e específico) – efeito retroativo geral é aquele produzido por um exame sem levar em conta suas características ou conteúdo; um exame que leva as pessoas a estudarem mais, independentemente de ser na direção dos conteúdos do exame, tem, portanto, efeito retroativo geral. Efeito retroativo específico, por outro lado, é definido como aquele que leva em conta o conteúdo do teste ou um aspecto específico desse conteúdo; b) *intensidade* (forte ou fraco) - o efeito retroativo pode ser forte ou fraco,

---

<sup>19</sup> *Backwash* é o termo predominante na literatura da área de educação, enquanto *washback* é mais comumente utilizado em estudos da área de linguística aplicada (de tradição inglesa) (ALDERSON; WALL, 1993, apud SCARAMUCCI, 2004).

dependendo da intensidade com que se manifesta na sala de aula. É forte quando determina praticamente todas as atividades de sala de aula ou todos os participantes; é fraco quando afeta parte da aula e alguns professores e alunos, mostrando a força de outros elementos; c) *extensão* (longo ou curto) - o efeito retroativo pode durar um período longo ou curto de tempo. Por exemplo, no caso de um exame vestibular, se os efeitos persistirem até os alunos entrarem em uma instituição, por exemplo, o efeito é curto; se persistirem após a entrada dos alunos, pode ser considerado longo. Essa dimensão pode ainda ser observada de uma outra perspectiva: um efeito que se observa logo após a introdução de um exame, mas que diminui com o tempo, é curto; e longo, por outro lado, quando esse efeito permanece por um longo tempo depois de o exame ter sido implementado; d) *intencionalidade* (intencional e não intencional): as consequências sociais intencionais e também as não intencionais de um exame são vistas como efeito retroativo (Messick, 1989; McNamara, 1996); e) *valor* (positivo e negativo): tem sido uma das dimensões mais discutidas, relacionada com intencionalidade — efeito retroativo positivo tem sido associado com efeito intencional (aquele desejado quando o exame é implementado) e efeito retroativo negativo (e positivo) com não intencional. Essa discussão é complexa por envolver julgamento de valor: um exame pode ser visto como positivo por professores e negativo por diretores, por exemplo. Segundo Watanabe (2004), esse julgamento deve ser feito sempre em relação aos candidatos ao teste (SCARAMUCCI, 2004, p. 206).

Dessa forma, nossas análises buscam entender os índices de confiabilidade, validade e as especificidades, intensidade, extensão, intencionalidade e valores do efeito retroativo dos testes e das práticas sociais aqui investigadas.

O próximo capítulo apresenta e detalha os procedimentos metodológicos adotados e implementados nesta pesquisa.

## CAPÍTULO 3 - METODOLOGIA

O presente estudo concentra-se na análise do TESLLE de Língua Inglesa e na seção de leitura (*reading comprehension*) do TOEFL ITP. No entanto, é importante destacar que as demais seções do TOEFL ITP (*listening comprehension* e *structure and written expression*) não são desconsideradas na presente pesquisa. Muito embora a análise textual tenha se concentrado na seção de leitura do TOEFL ITP, as seções de compreensão oral e de estrutura e expressão escrita são consideradas na discussão dos resultados da análise contextual, por serem parte integrante do teste TOEFL ITP.

Os procedimentos de análise implementados neste trabalho estão alinhados à perspectiva ética, ou seja, “cuja perspectiva adotada permite estudar a dinâmica social a partir do estudo do texto, como materialidade dessa dinâmica” (MOTTA-ROTH, 2003, p. 170). Isto é, nosso *corpus* (contextual e textual) é composto por textos que nos permitem investigar a prática social. Essa perspectiva se distingue da perspectiva êmica (de dentro), que busca “desvelar valores da cultura local por meio de entrevistas e observação detalhada junto aos membros dessas mesmas culturas” (MOTTA-ROTH, 2003, p. 172) e pode envolver análises etnográficas, por exemplo.

Para que se estabeleça uma Análise Crítica de Gênero do TESLLE e TOEFL ITP, buscamos entender esses testes de leitura de EAP como práticas sociais situadas, buscando estabelecer relações entre texto e contexto destes. Para isso, foram selecionados documentos e amostras do TESLLE e TOEFL ITP, de forma a investigar a estrutura e a função social desses testes em seus múltiplos contextos e especificidades. Tais contextos envolvem processos como, por exemplo, a regulamentação, elaboração, divulgação, aplicação e correção desses testes. A fim de realizar uma análise documental dessa prática social tão completa quanto possível, buscou-se entender a influência que esses diversos processos exercem sobre cada um dos testes. Dessa forma, foram exploradas as relações entre os contextos de produção, circulação e consumo (FAIRCLOUGH, 1992) dos testes a partir do *corpus*, que está descrito na seção a seguir.

### 3.1 O *CORPUS*: CRITÉRIOS DE SELEÇÃO E DESCRIÇÃO DA AMOSTRA

Todos os textos que compõem o *corpus* deste estudo foram obtidos dos *sites* das instituições responsáveis pela elaboração e/ou aplicação desses testes, ou de manuais oficiais que são indicados ou mesmo comercializados por essas instituições. O *corpus* deste estudo é apresentado pela Quadro 3, que também indica a equivalência entre os textos relacionados aos testes.

Quadro 3 - O *corpus* deste estudo

TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
<i>Corpus da análise contextual</i>	
<ul style="list-style-type: none"> <li>Resolução de n. 015/2014, que diz respeito ao regimento geral da pós-graduação stricto sensu e lato sensu da UFSM; e</li> <li>Resolução de n. 003/2010, que diz respeito ao TESLLE na UFSM e ao aproveitamento de testes de outras instituições.</li> </ul>	
<ul style="list-style-type: none"> <li><i>Site</i> do Teste de Suficiência em Língua Estrangeira da UFSM (ufsm.br /testedesuficiencia).</li> </ul>	<ul style="list-style-type: none"> <li><i>Site</i> da ETS, desenvolvedora dos testes TOEFL (ets.org/toefl_itp); e</li> <li><i>Site</i> da Mastertest, distribuidora do teste TOEFL ITP no Brasil (mastertest.com.br).</li> </ul>
<ul style="list-style-type: none"> <li>Mini-manual do candidato (TESLLE, 2018).</li> </ul>	<ul style="list-style-type: none"> <li>Manual do candidato (ETS, 2017).</li> </ul>
<ul style="list-style-type: none"> <li>Editais vigentes em 2019 e 2020 (TESLLE, 2021).</li> </ul>	<ul style="list-style-type: none"> <li><i>Official TOEFL Handbook</i> (ETS, 2012a);</li> <li><i>Supervisor's Manual</i> (ETS, 2012b);</li> <li><i>TOEFL ITP Test and Score Data Summary</i> (2019); e</li> <li><i>TOEFL ITP Test Score Descriptors</i> (2014b).</li> </ul>
<i>Corpus da análise textual</i>	
<p>Amostra do teste de Língua Inglesa (TESLLE, 2019).</p>	<p>Guia oficial para o TOEFL ITP (material didático que inclui uma amostra do teste) (ETS, 2014a).</p>

Fonte: elaborado pelo autor

Assim sendo, o presente *corpus* é composto por: i) duas resoluções da UFSM; ii) três *sites* institucionais; iii) dois manuais para os testes; iv) dois editais vigentes em 2019; v) um guia oficial; e vi) uma amostra de cada teste. A análise textual incumbiu-se mais especificamente das amostras dos testes. As duas amostras de teste analisadas neste estudo foram coletadas do *site* do TESLLE e de um guia oficial do TOEFL ITP. A amostra do TESLLE que compõe o *corpus* deste estudo estava disponível no *site* do TESLLE<sup>20</sup> durante o período da coleta do presente *corpus*, e a amostra do teste TOEFL ITP constitui o sexto capítulo do livro intitulado “Guia oficial para o teste TOEFL ITP” (ETS, 2014)<sup>21</sup>.

### 3.2 OS PROCEDIMENTOS DE ANÁLISE

Primeiramente, realizou-se a coleta do *corpus*, a partir dos *sites* oficiais das instituições que desenvolvem e distribuem os testes, conforme explicado na seção 3.1. Posteriormente, uma análise-piloto foi conduzida com a finalidade de identificar e entender os aspectos contextuais relevantes para a prática de testagem de leitura de EAP. Em seguida, foram feitas, simultaneamente, as análises dos textos e seus contextos, que estão descritas nas seções a seguir. Os dados encontrados foram submetidos a análises qualitativa e quantitativa. A abordagem quantitativa contribuiu para a identificação, por exemplo, do número de questões que se enquadram em cada tipo e enfoque e cada estrato da linguagem, enquanto que a abordagem qualitativa foi implementada na classificação das questões de acordo com seus tipos, enfoques, estratos, e na relação desses dados com os resultados da análise contextual.

#### 3.2.1 Análise contextual

A análise contextual voltou-se para a investigação dos contextos dos testes, que foi realizada a partir das informações dos documentos e *sites* oficiais do TOEFL ITP e do TESLLE (conforme demonstrado pelo Quadro 2, na seção 3.1). Essa análise teve a finalidade de identificar informações pertinentes para a contextualização e interpretação

---

<sup>20</sup> Endereço: [ufsm.br/pro-reitorias/prpgp/teslle/](http://ufsm.br/pro-reitorias/prpgp/teslle/). Acesso em: 29 de mar. 2019.

<sup>21</sup> Em inglês: *Official Guide to the TOEFL ITP Test*. Todas as referências a essa obra neste trabalho estão traduzidas pelo autor.

dos testes e(m) suas práticas sociais. Dessa forma, buscamos identificar como o *corpus* contextual apresenta e descreve: i) os objetivos sociocomunicativos; ii) o(s) público(s)-alvo; iii) o(s) gênero(s) ou tipos de texto(s)-base explorados; e iv) as questões dos testes. Essa análise possibilitou a identificação de diversos fatores que podem exercer influência nos processos da testagem, como as normas e condições de aplicação dos testes e a forma como esses testes são apresentados por suas respectivas instituições responsáveis.

### 3.2.2 Análise textual

Na análise textual, os textos-base dos testes foram analisados de forma a identificar e descrever seu(s) gênero(s), conteúdos e características estruturais, como a extensão, o *layout*, e a autenticidade.

Por sua vez, as questões dos testes foram analisadas e classificadas de acordo com: i) suas estruturas; ii) seus tipos; iii) os enfoques; e iv) os estratos e as competências que elas mobilizam<sup>22</sup>. Como referência para essa análise, partiu-se das categorias encontradas e utilizadas por Marcuzzo e Radünz (2017; 2019).

Para identificar o tipo de questão, investigou-se de que formas as questões desses testes estão estruturadas para estimular um raciocínio e solicitar uma resposta dos examinandos. Para identificar os enfoques, foram investigados marcadores linguísticos que manifestam um ou mais aspecto(s) ou habilidade(s) linguística(s) as questões abordam e demandam. Por fim, buscou-se identificar os estratos da linguagem que são centrais para a compreensão e o respondimento das questões. Essa análise seguiu as fundamentações de Halliday (1978) e Halliday e Hasan (1989), e utilizou os termos da estratificação da linguagem como sistema sociossemiótico (MARTIN, 1992; MOTTA-ROTH, 2008): Fonologia e Grafologia; Léxico-gramática; Semântica e Pragmática; Registro; Gênero; e Discurso (conforme ilustrado pela Figura 2 na seção 2.4).

A análise do texto incluiu ao todo 66 questões (16 do TESLLE de Língua Inglesa e 50 da seção de leitura do TOEFL ITP). Para identificação e posterior referência ao longo deste trabalho, as 66 questões do *corpus* receberam um código, o qual é composto pela

---

<sup>22</sup> Cada uma dessas categorias é explorada e compõe uma seção quaternárias na seção 4.1.2.

palavra TESLLE ou TOEFL, seguido da letra Q e de um número, separados por um hífen (-). As palavras TESLLE e TOEFL referem-se ao nome do teste, enquanto a letra Q e o número foram utilizados para se referir à palavra “questão” e ao número desta, respectivamente. Por exemplo: a primeira questão do TESLLE analisada está representada pelo código: TESLLE Q-01.

Por fim, o cruzamento dos resultados das análises do texto e do contexto possibilitou levantar discussões acerca das concepções de linguagem (HALLIDAY, 1978; HALLIDAY; MATTHIESSEN, 2014; THOMPSON, 2004) e letramento (STREET, 1984; ROJO, 2009), além de estimar a confiabilidade, a validade e o efeito retroativo potencial dos testes e da prática de testagem investigados (ALDERSON, CLAPHAM; WALL, 1995; SCARAMUCCI, 2004; 2009).

No capítulo 4 são apresentados e discutidos os resultados das análises realizadas neste estudo, buscando esclarecer as relações entre texto e contexto. Em um primeiro momento, são descritas as estruturas dos testes que compõem o *corpus*. Em seguida, são detalhados os achados acerca dos textos-base e das questões.

## **CAPÍTULO 4 - RESULTADOS E DISCUSSÃO**

Neste capítulo são descritos e discutidos os resultados das análises realizadas acerca do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP. Primeiramente, a seção 4.1 trata dos aspectos estruturais dos testes, isso é, de que forma esses testes são materializados e organizados linguisticamente, e em que medida suas estruturas refletem seus propósitos sociocomunicativos. Por sua vez, na seção 4.2, apontam-se e discutem-se os aspectos e as implicações sociais desses testes enquanto gêneros discursivos, com vistas a compreender seus processos de testagem em seus diversos contextos. Por fim, na seção 4.3 os resultados são sistematizados e interpretados.

### **4.1 A ESTRUTURA DOS TESTES**

O TESLLE de Língua Inglesa é composto por 16 questões enquadradas em uma única seção, sem subdivisões, além de um texto-base, como mostra a Figura 4.

Figura 4 - O TESLLE de Língua Inglesa

<p><b>Lingua Inglesa</b></p> <p>Para responder às questões, leia o texto a seguir.</p> <h2>Rigour, reliability and validity in qualitative research</h2> <p><b>T. Long and M. Johnson</b></p> <p>This article addresses issues relating to rigour within qualitative research, beginning with the need for rigour at all in such studies. The concept of reliability is then analysed, establishing the traditional understanding of the term, and evaluating alternative terms. A similar exploration of validity and proposed alternatives follows. It is suggested that there is nothing to be gained from the use of alternative terms which, on analysis, often prove to be identical to the traditional terms of reliability and validity. Alternative or novel means of addressing these concepts in interpretive research are, however, welcomed. A review of some of the strategies available for the pursuit of reliability and validity in qualitative research is undertaken. These are clearly identified as means to establish existing criteria and are found to have variable value. © 2000 Harcourt Publishers Ltd</p> <p><b>Keywords:</b> rigour, reliability, validity, qualitative research</p> <p><b>THE NEED FOR RIGOUR IN RESEARCH</b></p> <p>There is general agreement that all research studies must be open to critique and evaluation. Failure to assess the worth of a study – the soundness of its method, the accuracy of its findings, and the integrity of assumptions made or conclusions reached – could have dire consequences. Ambiguous or meaningless findings may result in wasted time and effort, while findings which are simply wrong could result in the adoption of dangerous or harmful practices. Evaluation of studies, then, is an essential prerequisite of the application of findings. Traditionally, such evaluation has centred on assessment of reliability and validity. However, while these terms have distinct meanings which relate well to other concepts and assumptions within the logical-positivist paradigm, their use in qualitative work has been questioned. A variety of positions is to be found. These include dismissing any attempt to establish rigour; using existing terms and criteria in a traditional manner; using existing terms and criteria with modification to their interpretation; and rejecting traditional terms and criteria while substituting new terms and criteria. It is the last of these which is challenged here.</p> <p><b>RELIABILITY</b></p> <p>Taking two commonly used sources for the traditional understanding of reliability, this concept can be described as “the consistency or constancy of a measuring instrument” (LoBiondo-Wood &amp; Haber 1998, p. 558), or “the degree of consistency or dependability with which an instrument measures the attribute it is designed to measure” (Polit &amp; Hungler 1995, p. 651). Hammersley (1992, p. 67) suggests that reliability “refers to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions.” Surprisingly, in view of their different philosophical approaches, there is little disagreement among these definitions. The first two indicate the detached nature of the researcher, while the third acknowledges the active involvement of the researcher, but all relate to confidence in data collection.</p> <p>The traditional understanding of reliability focuses on standardizing data collection instruments (Mason 1996, p. 24). However, this is premised on the assumption that methods of data generation can be conceptualised as tools, and can be standardized, neutral and non-biased” (Mason 1996, p. 145). While this may be acceptable for quantitative methods, though Hammersley (1992) questions even this, the non-standardization of qualitative methods and the determination to seek greater validity through retention of content makes it impossible in qualitative work. Brink (1991, p. 176) proposes three tests of reliability for qualitative work, each to be used as is appropriate for specific studies. Stability is established when asking identical questions of an informant at different times produces consistent answers. Consistency refers to the</p> <p><b>02</b></p>	<p><b>UFSM</b></p> <p>Responda às questões com base no texto intitulado <i>Rigour, reliability and validity in qualitative research</i>.</p> <p><b>01</b></p> <p>O tema do estudo reportado no texto é</p> <p><input type="radio"/> a) valorização da pesquisa qualitativa.</p> <p><input type="radio"/> b) rigor na pesquisa qualitativa.</p> <p><input type="radio"/> c) terminologia alternativa e rigor científico.</p> <p><input type="radio"/> d) rigor nas pesquisas em geral.</p> <p><input type="radio"/> e) qualidade e rigor na pesquisa científica.</p> <p><b>02</b></p> <p>No parágrafo 1, qual critério <b>NÃO</b> é citado para definir o valor de um estudo?</p> <p><input type="radio"/> a) Integridade das afirmações.</p> <p><input type="radio"/> b) Coerência dos objetivos.</p> <p><input type="radio"/> c) Solidez do método.</p> <p><input type="radio"/> d) Precisão dos resultados.</p> <p><input type="radio"/> e) Integridade das conclusões.</p> <p><b>03</b></p> <p>No segmento “It is the last of these which is challenged here” (§ 1), o segmento destacado se refere a</p> <p><input type="radio"/> a) “dismissing any attempt to establish rigour” (§ 1).</p> <p><input type="radio"/> b) “using existing terms and criteria in the traditional manner” (§ 1).</p> <p><input type="radio"/> c) “using existing terms and criteria with modification to their interpretation” (§ 1).</p> <p><input type="radio"/> d) “rejecting traditional terms and criteria while substituting new terms and criteria” (§ 1).</p> <p><input type="radio"/> e) “A variety of positions is to be found” (§ 1).</p> <p><b>04</b></p> <p>Nos parágrafos 2 e 3, os autores do texto apresentam várias definições para confiabilidade. Associe as colunas, relacionando a definição a seu(s) respectivo(s) autor(es).</p> <p>1. LoBiondo-Wood &amp; Haber (1998) <input type="checkbox"/> Grau de consistência ou dependência com que um instrumento mede um atributo.</p> <p>2. Polit &amp; Hungler (1995) <input type="checkbox"/> Grau de consistência com que ocorrências são atribuídas a mesma categoria por observadores diferentes ou pelo mesmo observador em situações distintas.</p> <p>3. Hammersley (1992) <input type="checkbox"/> Padronização dos instrumentos de coleta de dados.</p> <p>4. Mason (1996) <input type="checkbox"/> Consistência ou constância de um instrumento de medida.</p> <p>Assinale a sequência correta.</p> <p><input type="radio"/> a) 4 – 3 – 2 – 1.</p> <p><input type="radio"/> b) 4 – 1 – 3 – 2.</p> <p><input type="radio"/> c) 2 – 3 – 4 – 1.</p> <p><input type="radio"/> d) 3 – 2 – 1 – 4.</p> <p><input type="radio"/> e) 1 – 4 – 2 – 3.</p> <p><b>08</b></p>
--	--

Fonte: TESLLE (2019, p. 2 e 8).

O TESLLE é escrito em língua portuguesa, com exceção do texto-base e de trechos ou termos retirados deste que são mencionados em algumas das questões, que são escritos, então, em língua inglesa. De acordo com as informações disponibilizadas pelo site da UFSM<sup>23</sup>, as 16 questões que compõem esse teste são de múltipla-escolha, e cada uma possui cinco alternativas de resposta, já o texto-base não tem uma extensão definida. Além disso, o TESLLE de Língua Inglesa analisado tem uma extensão total de 12 páginas, nas quais, além da página da capa do teste, seis páginas contêm o texto-base e cinco páginas contêm as questões, e cada questão ocupa entre um sexto e meia página, sem exceder os limites da página em que se encontram. Ademais, esse teste tem duração máxima de 120 minutos, desconsiderando o tempo dedicado às instruções gerais fornecidas oralmente aos examinandos.

<sup>23</sup> Endereço: [ufsm.br/pro-reitorias/prpgp/teslle/](http://ufsm.br/pro-reitorias/prpgp/teslle/). Acesso em: 29 de mar. 2019.

Por sua vez, a versão ITP (*Level 2*) do TOEFL que foi analisada no presente estudo se distingue das demais modalidades de testes TOEFL (como do TOEFL iBT, por exemplo) por não ter seções dedicadas a aferir as habilidades em escrita (*writing*) ou produção oral (*speaking*). À vista disso, esse teste é dividido em três seções: compreensão oral (*listening comprehension*), que tem 50 questões e duração de 35 minutos; estrutura e expressão escrita (*structure and written expression*), que tem 40 questões e duração de 25 minutos; e leitura (*reading comprehension*), que tem 50 questões e duração de 55 minutos. Assim, o teste tem 140 questões e a sua realização tem duração de 1 hora e 55 minutos, desconsiderando o tempo dedicado às instruções gerais fornecidas oralmente e de preenchimento dos dados pessoais dos examinandos. Além disso, a seção de compreensão oral é ainda subdividida em três partes: Parte A: conversas breves (*Part A: short conversations*); Parte B: conversas estendidas (*Part B: extended conversations*); e Parte C: falas breves (*Part C: short talks*). Por sua vez, a seção de estrutura e expressão escrita subdivide-se em duas partes, em que a primeira (estrutura) “testa a habilidade de construir frases completas e gramaticalmente corretas”, e a segunda (expressão escrita) “testa a habilidade de detectar erros em escrita acadêmica” (ETS, 2014, p. 25). Por fim, a seção de leitura não é subdividida em termos de tipo de questão, muito embora o guia oficial do teste classifique-as em seis tipos: “ideia principal, informação factual, organização e lógica, relações referenciais, vocabulário em contexto, e inferência” (Ibid., p. 33). Além disso, os cinco textos-bases dessa seção são definidos como passagens curtas de texto entre 300 e 350 palavras (Ibid., p. 33).

A Figura 5 ilustra um dos textos-base e suas respectivas questões da seção de leitura do TOEFL ITP.

Figura 5 - A seção de leitura do TOEFL ITP<sup>24</sup>

---

<sup>24</sup> Note-se que a amostra da seção de leitura do TOEFL ITP aqui analisada, parte integrante do *Official guide*, está em dimensões maiores do que o teste original.

3	3	3	3	3	3	3	3	3	3	3	3	3	3
<p><b>Questions 1–7</b></p> <p>Hotels were among the earliest facilities that bound the United States together. They were both creatures and creators of communities, as well as symptoms of the frenetic quest for community. Even in the first part of the nineteenth century, Americans were already forming the habit of gathering from all corners of the nation for both public and private, business and pleasure, purposes. Conventions were the new occasions, and hotels were distinctively American facilities making conventions possible. The first national convention of a major party to choose a candidate for president (that of the National Republican Party, which met on December 12, 1831, and nominated Henry Clay for president) was held in Baltimore, at a hotel that was then reputed to be the best in the country. The presence in Baltimore of Barnum's City Hotel, a six-story building with two hundred apartments, helps explain why many other early national political conventions were held there.</p> <p>In the longer run, American hotels made other national conventions not only possible but pleasant and convivial. The growing custom of regularly assembling from afar the representatives of all kinds of groups—not only for political conventions, but also for commercial, professional, learned, and avocational ones—in turn supported the multiplying hotels. By the mid-twentieth century, conventions accounted for over a third of the yearly room occupancy of all hotels in the nation; about 18,000 different conventions were held annually with a total attendance of about ten million persons.</p> <p>Nineteenth-century American hotelkeepers, who were no longer the genial, deferential "hosts" of the eighteenth-century European inn, became leading citizens. Holding a large stake in the community, they exercised power to make it prosper. As owners or managers of the local "palace of the public," they were makers and shapers of a principal community attraction. Travelers from abroad were mildly shocked by this high social position.</p>							<p>1. The word "bound" in line 1 is closest in meaning to (A) led (B) protected (C) tied (D) strengthened</p> <p>2. The National Republican Party is mentioned in line 8 as an example of a group (A) from Baltimore (B) of learned people (C) owning a hotel (D) holding a convention</p> <p>3. The word "assembling" in line 14 is closest in meaning to (A) announcing (B) motivating (C) gathering (D) contracting</p> <p>4. The word "ones" in line 16 refers to (A) hotels (B) conventions (C) kinds (D) representatives</p> <p>5. The word "it" in line 22 refers to (A) European inn (B) host (C) community (D) public</p> <p>6. It can be inferred from the passage that early hotelkeepers in the United States were (A) active politicians (B) European immigrants (C) professional builders (D) influential citizens</p> <p>7. Which of the following statements about early American hotels is NOT mentioned in the passage? (A) Travelers from abroad did not enjoy staying in them. (B) Conventions were held in them. (C) People used them for both business and pleasure. (D) They were important to the community.</p>						
122	Official Guide to the TOEFL® ITP Test						Complete TOEFL ITP Practice Test						123

Fonte: *Official Guide to the TOEFL ITP* (2014, p. 122-123).

Assim, percebe-se que tanto no TESLLE quanto no TOEFL ITP há ocorrência de intertextualidade e interdiscursividade (FAIRCLOUGH, 1992, p. 85), no sentido em que há ocorrência explícita de outros textos nesses testes. O texto-base do TESLLE foi originalmente produzido para outra prática social, e foi recontextualizado para constituir o teste, o que também parece ser o caso dos textos-base da seção de leitura do TOEFL ITP, caracterizando, portanto, o que Fairclough entende por intertextualidade (manifesta) (Ibid., p. 85).

Também se constataram ocorrências de interdiscursividade (ou intertextualidade constitutiva), pois, nesses testes, há uma "constituição heterogênea de textos por meio de elementos (tipos de convenção) das ordens de discurso" (FAIRCLOUGH, 2001, p.

114). Isso acontece na medida em que esses testes seguem estruturas textuais convencionadas do gênero, ao, por exemplo, o(s) texto(s)-base preceder(em) as questões, as questões precederem as alternativas de resposta, etc.

Ao analisar a organização retórica desses testes, a partir da análise da forma e do conteúdo das partes que os compõem, identificamos sete movimentos, dos quais cinco são comuns a ambos os testes. Essa análise foi feita com base nas concepções propostas por Martin e Matthiessen (2012) e Swales (1990; 2004).

Quadro 4 – Organização retórica do gênero teste de leitura de EAP

(continua)

Movimento retórico		TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
1	<b>Capa</b>	Capa do livro teste	Capa do teste
2	<b>Instruções iniciais</b>	NC <sup>25</sup>	Instruções iniciais
3	<b>Exemplos</b>	NC	Texto-base
			Questões
			Respostas
			Explicações das respostas
4	<b>Comandos iniciais</b>	Comando inicial do teste	Comando inicial do teste
		Comando inicial do texto-base	Comando inicial do exemplo de texto-base
		Comando inicial das questões	Comando inicial das questões
5	<b>Texto-base</b>	Texto-base	Textos-base
6	<b>Questões</b>	Contextualização	NC
		Dados	NC
		Comandos	Comando

<sup>25</sup> Não consta na amostra analisada.

Quadro 4 – Movimentos Retóricos do gênero teste de leitura de EAP

(continuação)

Movimento retórico		TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
6	Questões	Alternativas	Alternativas
7	Instruções finais	NC	Instruções finais

Fonte: elaborado pelo autor.

A capa (movimento 1) é o primeiro elemento textual presente no TESLLE, ela indica a língua-alvo do teste (inglesa) e também a instituição e os departamentos envolvidos com o teste. Esse movimento retórico não está presente na amostra da seção de leitura do TOEFL ITP<sup>26</sup> analisada, pois essa amostra compõe uma seção do *Official Guide*. No entanto, a capa é presente e precede a seção de compreensão oral (*listening comprehension*) dos testes TOEFL ITP que são, de fato, aplicados<sup>27</sup>. Dessa forma, esta também é capa da seção de leitura do TOEFL ITP e, portanto, consideramos que esse movimento retórico está presente no gênero.

Em seguida, a seção de leitura do TOEFL ITP apresenta instruções iniciais (movimento 2), fornecendo orientações gerais escritas sobre como realizar o teste. Essas orientações podem ser feitas oralmente pelo aplicador do TESLLE, mas não pelo do TOEFL ITP, uma vez que a seção de leitura desse teste deve iniciar imediatamente após o término da seção anterior (*structure and written expression*), sem que haja qualquer “interrupção” por parte do aplicador do teste. A possibilidade dessas instruções serem dadas oralmente pelo aplicador do TESLLE, mas não do TOEFL, possivelmente é a causa da ausência dessas instruções escritas no TESLLE.

Esse movimento é seguido de um exemplo (movimento 3) de texto-base mais curto (de aproximadamente 100 palavras, enquanto os demais textos-base têm entre 300 e 350 palavras), duas questões correspondentes, que não são numeradas, dois exemplos

<sup>26</sup> É importante destacar que, neste estudo, entendemos a seção de leitura do TOEFL ITP não como um movimento retórico do gênero, mas como um gênero (teste) por si só.

<sup>27</sup> Essa informação não advém das amostras analisadas, mas do conhecimento prévio do autor, que aplica e realiza o teste TOEFL ITP regularmente.

de resposta dessas questões e uma breve explicação dessas respostas. Esse movimento retórico não está presente no TESLLE de Língua Inglesa, portanto, configura-se como não essencial para o gênero teste de leitura de EAP.

Há também, em ambos os testes, um comando inicial do teste (movimento 4), que orienta os examinandos a responder às questões (movimento 6) com base em seus respectivos textos-base (movimento 5). É importante que se faça a distinção entre as instruções iniciais (movimento 2) e o comando inicial do teste (movimento 4). Nas instruções iniciais, presentes apenas na seção de leitura do TOEFL ITP, são apresentadas informações gerais sobre a seção

Nesta seção você irá ler várias passagens. Cada uma é seguida por algumas questões sobre ela. Escolha a melhor resposta — (A), (B), (C) ou (D) — para cada questão. E então, na sua folha-resposta, encontre o número da questão e preencha o espaço que corresponde à letra da resposta que você escolheu (ETS, 2014, p. 119).

Já os comandos iniciais dos testes (movimento 4), servem para orientar os examinandos especificamente sobre o que se espera que eles façam. Na seção de leitura do TOEFL ITP, foram encontrados três comandos iniciais, dois antes do exemplo (movimento 3): “Responda todas as perguntas que seguem a passagem com base no que está afirmado ou implícito na passagem” e “Leia a passagem a seguir”; e um logo após os exemplos, precedendo o primeiro texto-base: “Agora comece o trabalho nas questões” (ETS, 2014, p. 119-120). No TESLLE de Língua Inglesa foram identificados dois comandos iniciais, precedendo, respectivamente, o texto-base a as questões do teste: “Para responder às questões, leia o texto a seguir” (TESLLE, p. 2), e “Responda às questões com base no texto intitulado *Rigour, reliability and validity in qualitative research*” (Ibid., p. 8).

As questões (movimento 6) dos testes estão constituídas por um comando e alternativas de resposta. As questões do TESLLE podem também apresentar uma breve contextualização seguida por dados no enunciado da questão. Essas características das questões de ambos os testes são descritas discutidas de forma mais aprofundada na seção 4.1.2.2.

As instruções finais (movimento 7) também se configuram como um movimento retórico opcional, encontrado apenas no fim da seção de leitura do TOEFL ITP. Essas instruções sinalizam o fim do teste e orientam o examinando a revisar suas respostas dentro do tempo limite para a realização da seção.

Assim, constatamos que quatro dos sete movimentos retóricos do gênero encontrados na amostra analisada estão presentes em apenas um dos testes analisados neste estudo, e três deles estão presentes apenas na seção de leitura do TOEFL ITP. Os movimentos retóricos que não são comuns a ambos os testes são a capa (movimento 1), que só foi encontrada na amostra do TESLLE, enquanto os movimentos retóricos de instruções iniciais (movimento 2), exemplos (movimento 3) e instruções finais (movimento 7) foram identificados apenas na seção de leitura do TOEFL ITP. No entanto, levando em conta que a capa é um elemento presente nos testes TOEFL ITP que são aplicados, podemos considerar, a partir da análise das duas amostras do *corpus*, que apenas três movimentos retóricos (2, 3 e 7) parecem ser, de fato, movimentos retóricos não essenciais ao gênero. Há, ainda, dentro do movimento retórico das questões (movimento 6), dois passos retóricos (contextualização e dados) que foram identificados apenas no TESLLE de Língua Inglesa.

Nesse sentido, de maneira geral, pode-se concluir que a seção de leitura do TOEFL ITP é um gênero mais complexo que o TESLLE de Língua Inglesa, por apresentar um maior número de movimentos retóricos na amostra analisada. Ao mesmo tempo, o TESLLE de Língua Inglesa apresenta questões estruturalmente mais complexas (com mais movimentos retóricos) que a seção de leitura do TOEFL ITP. Essa discussão é retomada a partir dos resultados das análises dos textos-base e as questões desses testes, que são detalhadas nas seções 4.1.1 e 4.1.2.

#### **4.1.1 Os textos-base**

O Quadro 5 foi elaborado de forma a traçar um paralelo entre como são descritos os aspectos gerais dos textos base do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP. Os resultados apresentados no Quadro 5 foram obtidos na análise contextual, apresentada pelo Quadro 3, na seção 3.1.

Quadro 5 – Descrições dos textos-base do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP.

Descrições dos textos-base	
TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
<p>“Um texto-base, sem extensão definida”</p> <p>“textos de publicações científicas e acadêmicas”</p> <p>“Fragmentos textuais de publicações de cunho científico e acadêmico.” (UFSM<sup>28</sup>, 2019)</p>	<p>Cinco passagens de 300-350 palavras (ETS, 2014, p. 33)</p> <p>“passagens de leitura” (Ibid., p. 2)</p> <p>“Material de leitura acadêmica escrito em inglês” (Ibid., p. 150)</p> <p>“Contexto e informações suficientes são fornecidas em cada passagem para que você possa responder às questões sem ter que recorrer a conhecimento de assuntos específicos fora da passagem.” (Ibid., p. 33)</p> <p>“uma variedade de passagens curtas sobre assuntos acadêmicos” (ETS, 2020b).</p>

Fonte: elaborado pelo autor.

O texto-base do TESLLE é descrito como um texto ou fragmentos textuais de publicações de cunho científico e acadêmico (UFSM, 2019), descrição que também não especifica suficientemente os possíveis gêneros dos textos-base do teste, além de indicar a possível ocorrência de recortes no texto original. Já as características dos textos-base do TOEFL ITP são descritas pelo guia oficial e pelo *site* apenas como passagens breves

<sup>28</sup> Endereço: [ufsm.br/pro-reitorias/prpgp/teslle/](http://ufsm.br/pro-reitorias/prpgp/teslle/). Acesso em: 29 de mar. 2019.

de leitura de tópicos acadêmicos variados em nível introdutório, similares em assunto e estilo àqueles lidos em cursos ensinados em instituições acadêmicas dos Estados Unidos (ETS, 2014; 2020b). Além disso, a ETS afirma que “contexto e informações suficientes são fornecidas em cada passagem para que você possa responder às questões sem ter que recorrer a conhecimento de assuntos específicos fora da passagem” (ETS, 2014, p. 33). Dessa forma, essa análise demonstrou haver uma aparente despreocupação em especificar o(s) gênero(s) dos textos-base, uma vez que uma infinidade de gêneros faz parte do âmbito acadêmico, e a descrição concentra-se nos possíveis assuntos, na extensão e no aspecto “autoexplicativo” das passagens. Ademais, a autoria e fonte dos textos-base não é mencionada no *corpus* contextual.

A análise do texto-base do TESLLE de Língua Inglesa indicou que se trata de um artigo acadêmico completo, de aproximadamente cinco mil palavras, que pode ser definido como um texto autêntico, ou seja, que foi escrito “para cumprir um propósito social na comunidade linguística” (LITTLE; DEVITT; SINGLETON, 1989, p. 25).

Figura 6 - Texto-base do TESLLE de Língua Inglesa

Lingua Inglesa

Para responder às questões, leia o texto a seguir.

## Rigour, reliability and validity in qualitative research

T. Long and M. Johnson

This article addresses issues relating to rigour within qualitative research, beginning with the need for rigour at all in such studies. The concept of reliability is then analysed, establishing the traditional understanding of the term, and evaluating alternative terms. A similar exploration of validity and proposed alternatives follows. It is suggested that there is nothing to be gained from the use of alternative terms which, on analysis, often prove to be identical to the traditional terms of reliability and validity. Alternative or novel means of addressing these concepts in interpretive research are, however, welcomed. A review of some of the strategies available for the pursuit of reliability and validity in qualitative research is undertaken. These are clearly identified as means to establish existing criteria and are found to have variable value. © 2000 Harcourt Publishers Ltd

**Keywords:** rigour, reliability, validity, qualitative research

### THE NEED FOR RIGOUR IN RESEARCH

There is general agreement that all research studies must be open to critique and evaluation. Failure to assess the worth of a study – the soundness of its method, the accuracy of its findings, and the integrity of assumptions made or conclusions reached – could have dire consequences. Ambiguous or meaningless findings may result in wasted time and effort, while findings which are simply wrong could result in the adoption of dangerous or harmful practices. Evaluation of studies, then, is an essential pre-requisite of the application of findings. Traditionally, such evaluation has centred on assessment of reliability and validity. However, while these terms have distinct meanings which relate well to other concepts and assumptions within the logical-positivist paradigm, their use in qualitative work has been questioned. A variety of positions is to be found. These include dismissing any attempt to establish rigour, using existing terms and criteria in the traditional manner, using existing terms and criteria with modification to their interpretation; and rejecting traditional terms and criteria while substituting new terms and criteria. It is the last of these which is challenged here.

### RELIABILITY

Taking two commonly used sources for the traditional understanding of reliability, this concept can be described as ‘the consistency or constancy of a measuring instrument’ (Lobbin-Wood & Haber 1978, p. 558), or ‘the degree of consistency or dependability with which an instrument measures the attribute it is designed to measure’ (Polit & Hungler 1995, p. 651). Hammersley (1992, p. 67) suggests that reliability ‘refers to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions.’ Surprisingly, in view of their different philosophical approaches, there is little disagreement among these definitions. The first two indicate the detached nature of the researcher, while the third acknowledges the active involvement of the researcher, but all relate to confidence in data collection.

The traditional understanding of reliability focuses on standardizing data collection instruments (Mason 1996, p. 24). However, ‘this is premised on the assumption that methods of data generation can be conceptualised as tools, and can be standardised, neutral and non-biased’ (Mason 1996, p. 145). While this may be acceptable for quantitative methods, though Hammersley (1992) questions even this, the non-standardization of qualitative methods and the determination to seek greater validity through retention of context makes it impossible in qualitative work. Brink (1991, p. 176) proposes three tests of reliability for qualitative work, each to be used as is appropriate for specific studies. *Stability* is established when asking identical questions of an informant at different times produces consistent answers. *Consistency* refers to the

§ 1
§ 2
§ 3

**Tony Long** PhD, BEd  
Senior Lecturer,  
School of Healthcare,  
Health, Behavior and Society,  
University of Leeds, PC,  
Beaumont, Leeds LS2 9JT,  
UK

**Martin Johnson** PhD, BEd  
Professor of Nursing, Dept  
of Adult & Critical Care  
Nursing, General  
Building, University of  
Central Lancashire, Preston  
PR1 2HE, UK

Correspondence to: T. Long,  
University of Leeds, Ltd,  
+44 (0) 113 233 2373 (ext. 400)  
+44 (0) 113 233 2374  
E-mail: a.lj@leeds.ac.uk

© 2000 Harcourt Publishers Ltd

Fonte: TESLLE, 2019, p. 8-9.

Nossa análise também confirma que o texto-base do TESLLE geralmente “não sofre alterações de *layout* nem recortes a fim de manter o texto no seu formato original” (MARCUSO; RADÜNZ, 2019, p. 3647). As únicas alterações encontradas na forma como o texto original é apresentado pelo TESLLE como texto-base foram: indicação de parágrafos (às margens do texto, separados por caixas); marca d’água em transparência atrás do texto-base (e também das questões); diagramação de cabeçalho e rodapé, para fins de padronização do *layout* do teste; e um recorte na lista de referências do texto original, que está indicado pela marcação “[...]”, feito para otimizar o enquadramento do texto-base nas páginas do teste (TESLLE, 2019, p. 7)<sup>29</sup>. Essas alterações parecem tentar

<sup>29</sup> Essas informações foram obtidas a partir de um ensaio realizado pelo autor submetido para a disciplina PPGLET824 Multimodalidade, Multimedialidade e Sociedade Contemporânea, do Programa de Pós-Graduação em Letras, UFSM, sob supervisão da professora Graciela Rabuske Hedges. Esse estudo, que foi submetido para publicação, apresenta uma análise do sistema composicional que investigou em que medida os elementos visuais de uma amostra do TESLLE de Língua Inglesa são articulados para otimizar a realização do teste.

não “interferir” nos significados do texto, ocorrendo, sempre que possível, nas margens deste e de forma discriminada.

Percebe-se, portanto, que há uma preocupação por parte dos elaboradores do TESLLE em manter a autenticidade (e o *layout*) do texto-base. Isso pode ser atribuído à perspectiva de linguagem subjacente ao TESLLE, explorada pelo trabalho de Marcuzzo e Radünz (2019), que afirma que a “perspectiva de linguagem que subjaz ao TESLLE é a “linguagem como gênero”, uma vez que o texto-base das questões é apresentado da mesma forma como circula nas práticas sociais, ou seja, com a mesma extensão e o mesmo *layout*. (Ibid., p. 3652). Essa perspectiva de linguagem zela não só pela autenticidade, mas também pela “integridade” visual do gênero, de forma que não há alteração nos significados (verbais e visuais) do texto quando utilizado para os fins da testagem. Não alterar as informações do texto-base também pode estar relacionado com motivos acadêmicos e éticos, pois, ao manter todos os aspectos do texto-base conforme sua publicação original, há um comprometimento com a confiabilidade das informações científicas e a propriedade intelectual dos autores.

O tema do texto-base da amostra analisada trata de rigor em pesquisas qualitativas, o que também confirma a afirmação de Marcuzzo e Radünz de que o tema do texto-base do TESLLE “é geralmente comum a todas as áreas [da ciência], como metodologia de pesquisa, plágio no contexto acadêmico, popularização do conhecimento científico, etc.” (Ibid., p. 3647).

A Figura 7 apresenta os dois primeiros textos-base da amostra da seção de leitura do TOEFL ITP analisada.

Figura 7 - Textos-base da seção de leitura do TOEFL ITP

	
<b>Questions 1–7</b>	<b>Questions 8–15</b>
<p>Hotels were among the earliest facilities that bound the United States together. They were both creatures and creators of communities, as well as symptoms of the frenetic quest for community. Even in the first part of the nineteenth century, Americans were already forming the habit of gathering from all corners of the nation for both public and private, business and pleasure, purposes. Conventions were the new occasions, and hotels were distinctively American facilities making conventions possible. The first national convention of a major party to choose a candidate for president (that of the National Republican Party, which met on December 12, 1831, and nominated Henry Clay for president) was held in Baltimore, at a hotel that was then reputed to be the best in the country. The presence in Baltimore of Barnum’s City Hotel, a six-story building with two hundred apartments, helps explain why many other early national political conventions were held there.</p> <p>In the longer run, American hotels made other national conventions not only possible but pleasant and convivial. The growing custom of regularly assembling from afar the representatives of all kinds of groups—not only for political conventions, but also for commercial, professional, learned, and avocational ones—in turn supported the multiplying hotels. By the mid-twentieth century, conventions accounted for over a third of the yearly room occupancy of all hotels in the nation; about 18,000 different conventions were held annually with a total attendance of about ten million persons.</p> <p>Nineteenth-century American hotelkeepers, who were no longer the genial, deferential “hosts” of the eighteenth-century European inn, became leading citizens. Holding a large stake in the community, they exercised power to make it prosper. As owners or managers of the local “palace of the public,” they were makers and shapers of a principal community attraction. Travelers from abroad were mildly shocked by this high social position.</p>	<p>With Robert Laurent and William Zorach, direct carving enters into the story of modern sculpture in the United States. Direct carving—in which the sculptors themselves carve stone or wood with mallet and chisel—must be recognized as something more than just a technique. Implicit in it is an aesthetic principle as well: that the medium has certain qualities of beauty and expressiveness with which sculptors must bring their own aesthetic sensibilities into harmony. For example, sometimes the shape or veining in a piece of stone or wood suggests, perhaps even dictates, not only the ultimate form, but even the subject matter.</p> <p>The technique of direct carving was a break with the nineteenth-century tradition in which the making of a clay model was considered the creative act and the work was then turned over to studio assistants to be cast in plaster or bronze or carved in marble. Neoclassical sculptors seldom held a mallet or chisel in their own hands, readily conceding that the assistants they employed were far better than they were at carving the finished marble.</p> <p>With the turn-of-the-century Arts and Crafts movement and the discovery of nontraditional sources of inspiration, such as wooden African figures and masks, there arose a new urge for hands-on, personal execution of art and an interaction with the medium. Even as early as the 1880s and 1890s, nonconformist European artists were attempting direct carving. By the second decade of the twentieth century, Americans—Laurent and Zorach most notably—had adopted it as their primary means of working.</p> <p>Born in France, Robert Laurent (1890–1970) was a prodigy who received his education in the United States. In 1905 he was sent to Paris as an apprentice to an art dealer, and in the years that followed he witnessed the birth of Cubism, discovered primitive art, and learned the techniques of woodcarving from a frame maker.</p> <p>Back in New York City by 1910, Laurent began carving pieces such as <i>The Priestess</i>, which reveals his fascination with African, pre-Columbian, and South Pacific art. Taking a walnut plank, the sculptor carved the expressive, stylized design. It is one of the earliest examples of direct carving in American sculpture. The plank’s form dictated the rigidly frontal view and the low relief. Even its irregular shape must have appealed to Laurent as a break with a long-standing tradition that required a sculptor to work within a perfect rectangle or square.</p>
<hr/> 122	<hr/> 124
Official Guide to the TOEFL® ITP Test	Official Guide to the TOEFL® ITP Test

Fonte: ETS, 2014, p. 122 e 124.

A análise das amostras dos testes indicou que os textos-base da seção de leitura do TOEFL ITP se assemelham a textos de popularização da ciência, bem como aponta Kamijo (2009, p. 204), ao afirmar que esses textos-base são de tipo histórico/narrativo, afirmação que parece estar alinhada com a perspectiva da linguística textual.

Os textos-base dessa seção também são descritos pelo *site* da ETS como “uma variedade de passagens curtas sobre assuntos acadêmicos” (ETS, 2020b). Esses assuntos acadêmicos podem tratar de:

**Artes:** belas artes, artesanato, teatro, dança, arquitetura, literatura, música, cinema, fotografia; **Humanidades:** história, ciência política, governamental, filosofia, direito; **Ciências da vida:** paleontologia, bioquímica, comportamento animal, ecologia, anatomia, fisiologia, genética, ciências da saúde, biologia, agricultura; **Ciências físicas:** geologia,

astronomia, química, ciências da terra, engenharia, meteorologia, energia, tecnologia, oceanografia, física; **Ciências sociais:** antropologia, sociologia, educação, geografia, arqueologia, psicologia, economia, negócios, administração, *marketing*, comunicações (ETS, 2020a, grifo nosso).

No caso da amostra da seção de leitura do TOEFL ITP analisada, os cinco textos-base parecem se enquadrar nessa descrição. Eles tratam, respectivamente, dos temas história, belas artes (escultura), comportamento animal, literatura (sátira) e astronomia.

Além disso, os textos-base não apresentam marcas do *layout* original ou qualquer indicação de aspectos contextuais como autoria, fontes e data de publicação, o que parece confirmar as constatações de Marcuzzo e Azambuja (2017, p. 347), de que os textos-base da seção de leitura do TOEFL ITP “[...] parecem ser textos não autênticos [...] criados para fins específicos dessa testagem”. Um fator importante a ser levado em consideração, no entanto, é que os testes TOEFL circulam em escala global, o que provavelmente dificulta o uso de textos autênticos como textos-base nos testes dadas as leis, normas e restrições sobre direitos autorais de cada país. À vista disso, é possível que seja mais viável a elaboração de textos próprios para os testes TOEFL do que a negociação de direitos autorais e a adaptação às regulamentações dos diversos países em que esses testes circulam.

#### 4.1.2 As questões

Essa seção terciária está organizada e subdividida em cinco seções quaternárias, que tratam dos aspectos contextuais (4.1.2.1), das estruturas (4.1.2.2), dos tipos (4.1.2.3), dos enfoques (4.1.2.4), e dos estratos e competências (4.1.2.5) das questões dos testes analisados neste estudo, respectivamente.

##### 4.1.2.1 Aspectos contextuais das questões

O Quadro 6, bem como o Quadro 5 da seção anterior, foi elaborado de forma a ilustrar em paralelo os aspectos textuais do TESLLE e da seção de leitura do TOEFL ITP, concentrando-se nas questões destes. Os resultados apresentados no Quadro 6 também

foram obtidos a partir da análise-piloto, que investigou o *corpus* contextual apresentado pelo Quadro 3, na seção 3.1, e com informações de estudos prévios.

Quadro 6 - Descrições das questões do TESLLE e da seção de leitura do TOEFL ITP

Aspectos textuais	TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
Estrutura das questões	<b>Tipos de questão</b>	<p>“Todas as [50] questões são de múltipla-escolha” (ETS, 2020b, tradução nossa)<sup>31</sup></p> <p>Normalmente há de nove a 11 questões por passagem (ETS, 2014, p. 33)</p>
	<b>Enfoques de questão</b>	<p>“Uma parte das [16] questões trata da compreensão do texto e habilidades de leituras e a outra parte é de questões léxico-gramaticais” (UFSM, 2019)</p> <p>“ideia principal, informação factual, organização e lógica, relações referenciais, vocabulário em contexto, e inferência.” (ETS, 2014, p. 33)</p>

Fonte: elaborado pelo autor.

A presente análise demonstrou que os tipos de questões de ambos os testes são pouco descritos pelo *corpus* contextual. Em contrapartida, os tipos de questão já foram investigados por autoras como Becker (2013), Cunha (2015), Marcuzzo e Radünz (2017; 2019) que se ocuparam de análises acerca do TESLLE, e por estudos como o de Perez-Amurao (2011) e Marcuzzo e Azambuja (2017), que, dentre outros fatores, se ocuparam de analisar as questões da seção de leitura do TOEFL ITP.

#### 4.1.2.2 As estruturas das questões

<sup>30</sup> Questões objetivas, nesse caso, se opõem a questões dissertativas. Nessa concepção, um dos tipos de questão objetiva é o de múltipla-escolha.

<sup>31</sup> Em inglês: *All questions are multiple choice*. Disponível em: [https://www.ets.org/toefl\\_itp/faq/](https://www.ets.org/toefl_itp/faq/). Acesso em: 27 de jun. 2020.

A Figura 8 foi elaborada de forma a ilustrar, a partir de dois exemplos de questão, de que forma as questões dos testes aqui investigados se estruturam com relação aos passos retóricos (tipos de informação) que apresentam.

Figura 8 - Estrutura das questões do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP

TESLLE de Língua Inglesa		Seção de leitura do TOEFL ITP	
<b>Contextualização</b>	<b>04</b> Nos parágrafos 2 e 3, os autores do texto apresentam várias definições para confiabilidade. Associe as colunas, relacionando a definição a seu(s) respectivo(s) autor(es).	<b>Comando</b>	32. According to the passage, there is a need for satire because people need to be
<b>Comando</b>	1. LoBiondo-Wood & Haber (1998) ( ) Grau de consistência ou dependência com que um instrumento mede um atributo. 2. Polit & Hungler (1995) 3. Hammersley (1992) ( ) Grau de consistência com que ocorrências são atribuídas a mesma categoria por observadores diferentes ou pelo mesmo observador em situações distintas. 4. Mason (1996)	<b>Alternativas</b>	(A) informed about new scientific developments (B) exposed to original philosophies when they are formulated (C) reminded that popular ideas are often inaccurate (D) told how they can be of service to their communities
<b>Dados</b>	( ) Padronização dos instrumentos de coleta de dados. ( ) Consistência ou constância de um instrumento de medida.		
<b>Comando</b>	Assinale a sequência correta.		
<b>Alternativas</b>	(a) 4 - 3 - 2 - 1. (b) 4 - 1 - 3 - 2. (c) 2 - 3 - 4 - 1. (d) 3 - 2 - 1 - 4.		

Fonte: TESLLE, 2019, p. 8; ETS, 2014, p. 131.

Na análise textual foi constatado que as questões do TESLLE de Língua Inglesa podem apresentar mais passos em seus enunciados do que a seção de leitura do TOEFL ITP, como demonstra a questão TESLLE Q-04, ilustrada na Figura 8, cujo enunciado é composto por uma breve contextualização, seguida por um primeiro comando, dados (a serem associados), um segundo comando e, por fim, as alternativas de resposta. Note-se que os passos entre parêntese na Figura 8 não ocorrem em todas as questões do teste. Mais especificamente, 43.75% (sete ocorrências) das questões do TESLLE de

Língua Inglesa apresentam mais de dois passos retóricos em suas estruturas, como mostra o Quadro 7.

Quadro 7 - Estrutura das questões do TESLLE de Língua Inglesa e da seção de leitura do TOEFL ITP

Nº	%	Questões do TESLLE de Língua Inglesa	Passos
09	56.25%	TESLLE Q-01, TESLLE Q-07, TESLLE Q-08, TESLLE Q-09, TESLLE Q-10, TESLLE Q-11, TESLLE Q-13, TESLLE Q-14 e TESLLE Q-16	02
03	18.75%	TESLLE Q-05, TESLLE Q-12 e TESLLE Q-15	04
02	12.5%	TESLLE Q-02 e TESLLE Q-03	03
02	12.5%	TESLLE Q-04 e TESLLE Q-06	05

Fonte: Elaborado pelo autor

Esses dados demonstram um considerável índice de questões padronizadas, como as da seção de leitura do TOEFL ITP, em que todas questões são compostas apenas por um comando e quatro alternativas de resposta, ou seja, dois passos retóricos. Isso também demonstra uma maior variedade e complexidade na formulação das questões do TESLLE de Língua Inglesa em relação à seção de leitura do TOEFL ITP.

#### 4.1.2.3 Tipos de questão

As 16 questões do TESLLE de Língua Inglesa são de múltipla-escolha, e, de acordo com Marcuzzo e Radünz (2019, p. 3617), estruturam-se em três tipos: pergunta e resposta; acordo ou desacordo; e associação. Ao analisar a amostra do TESLLE, foi possível identificar outros dois tipos de questão: pergunta indireta e imperativa. É importante notar que esses tipos de questão não estão presentes exclusivamente na amostra aqui analisada, mas que nossa classificação busca aprofundar as categorias identificadas pelos estudos de Marcuzzo e Radünz (2017; 2019).

Na categoria acordo ou desacordo, as questões demandam que o examinando identifique as alternativas ou afirmações/segmentos que estão em acordo ou desacordo

com as informações do texto-base. Em duas das três ocorrências de questões desse tipo na amostra analisada, também é necessário que o examinando assinale uma alternativa que indica a sequência de resposta correta dos segmentos postos pelas questões (TESLLE Q-05 e TESLLE Q-06). Nesses casos, essas questões são classificadas como de tipo acordo ou desacordo, pois o primeiro comando delas demanda que essa classificação seja indicada nos segmentos (onde há espaços designados para a indicação). Só em um segundo momento deve-se associar a sequência de classificações com as alternativas, e então assinalar uma destas. Portanto as questões TESLLE Q-05 e TESLLE Q-06 foram classificadas como de tipo acordo ou desacordo.

Definimos as questões de associação conforme Medeiros (1977, p. 79), que afirma que elas apresentam dois grupos de informação que precisam ser associados. Além disso, esse tipo de questão também demanda que uma alternativa com a sequência de resposta correta seja assinalada. As duas questões desse tipo encontradas no *corpus* têm um número igual de itens a serem relacionados, na contramão do que sugerem manuais para a elaboração de questões de testes, como, por exemplo, as diretrizes para elaboração e revisão de questões em processos seletivos elaboradas para o contexto da UFSM<sup>32</sup> (FUZER et al., 2017), que afirmam que “é aconselhável que as colunas apresentem um número diferente de itens, a fim de minimizar o acerto pela simples correspondência” (FUZER et al., 2017, p. 56).

Por fim, as questões imperativas<sup>33</sup>, como o nome sugere, apresentam todo enunciado no modo imperativo, com um ponto final. É importante notar que algumas questões de tipo acordo ou desacordo também apresentam seus enunciados no modo imperativo. No entanto, para classificar os tipos de questão, levamos em consideração o que a questão demanda que o examinando faça. Nas questões imperativas, é solicitado que o examinando assinale uma alternativa de acordo com determinado critério, ao passo que, nas de associação, é solicitado que se relacionem informações que precedem as alternativas.

---

<sup>32</sup> Material utilizado para a capacitação de bancas elaboradoras e revisoras de testes e processos seletivos da UFSM, que se constitui como um guia para a elaboração responsável, harmônica, uniforme e padronizada destes (FUZER et al., 2017, p.15).

<sup>33</sup> Esse tipo de questão foi classificado como de pergunta e resposta por Marcuzzo e Radünz (2017; 2019).

Quadro 8 - Tipos de questão do TESLLE de Língua Inglesa

TESLLE de Língua Inglesa			
Tipo	N <sup>o34</sup>	%	Excerto de questão
Pergunta indireta	06	37.5%	“O tema do estudo reportado no texto é” (TESLLE Q-01)
Pergunta e resposta	03	18.75%	“No parágrafo 1, qual critério <u>NÃO</u> é citado para definir o valor de um estudo?” (TESLLE Q-02)
Acordo ou desacordo	03	18.75%	“Sobre a confiabilidade, indique se os segmentos a seguir estão em (A) ou (D) com acordo ou desacordo o que apresenta o parágrafo 3.” (TESLLE Q-05)
Associação	02	12.5%	“Nos parágrafos 2 e 3, os autores do texto apresentam várias definições para confiabilidade. Associe as colunas, relacionando a definição a seu(s) respectivo(s) autor(es).” (TESLLE Q-04)
Imperativa	02	12.5%	“Assinale a alternativa cujo fragmento sublinhado indica alto grau de certeza do autor em relação ao que afirma.” (TESLLE Q-07)

Fonte: elaborado pelo autor.

Como mostra o Quadro 8, o TESLLE de Língua Inglesa apresenta cinco tipos de questão, em que uma parcela expressiva é de tipo pergunta indireta (37.5%), pergunta e resposta (18.75%) e de acordo ou desacordo (18.75%). No entanto, esses resultados percentuais foram obtidos a partir da análise de apenas um teste, o que é uma amostra muito pequena. Portanto, nos limitamos a fazer meras especulações com base nesse resultado e em nosso conhecimento prévio como participantes dessa prática social. Análises de um *corpus* mais expressivo são necessárias para explicar as razões pelas quais as questões de “associação” e “imperativa” são menos comuns no *corpus* deste estudo. Além disso, as questões de tipo imperativa parecem ser apenas uma formulação

<sup>34</sup> Nos quadros deste trabalho, N<sup>o</sup> refere-se ao número de ocorrências.

diferente do enunciado de questões como as de tipo pergunta e resposta e pergunta indireta. Outras pesquisas podem investigar as causas e impactos dessas diferentes formulações de questão, que não são o foco do presente estudo.

A seção de leitura do TOEFL ITP é descrita como tendo 50 questões de tipo múltipla-escolha, distribuídas em cinco passagens, que normalmente têm de nove a 11 questões correspondentes cada. No entanto, a análise do texto do teste apontou que há textos-base com sete, oito e até 12 questões correspondentes. Portanto, a afirmação de que cada texto-base “normalmente há de nove a 11 questões por passagem” (ETS, 2014, p. 33) não se aplica à amostra analisada, uma vez que as questões de quatro dos cinco textos-base não se enquadram nessa “normalidade”.

Além disso, todas as questões do TOEFL ITP são de múltipla escolha e podem ser subsequentemente classificadas em dois tipos: pergunta indireta; e pergunta e resposta (ou pergunta direta). Na pergunta indireta, o enunciado da questão se completa nas alternativas de resposta, e termina sempre sem pontuação ou com dois pontos, nunca com ponto de interrogação. Esse é o tipo de questão mais comum na seção de leitura do TOEFL ITP, como demonstra o Quadro 9. A categoria pergunta e resposta se trata de pergunta direta, devidamente marcada com ponto de interrogação. No TOEFL ITP, todas as questões desse tipo são também *WH questions*<sup>35</sup>.

Quadro 9 - Tipos de questão da seção de leitura do TOEFL ITP

Seção de leitura do TOEFL ITP			
Tipo	Nº	% <sup>36</sup>	Excerto de questão
Pergunta indireta	33	66%	“The word “bound” in line 1 is close in meaning to” (TOEFL Q-01)
Pergunta e resposta	17	34%	“Which of the following statements about early American hotels is NOT mentioned in the passage?” (TOEFL Q-07)

Fonte: elaborado pelo autor.

<sup>35</sup> *WH question* é um termo comumente usado para se referir a perguntas formuladas a partir de pronomes ou advérbios interrogativos que são o ponto de partida da pergunta. São pronomes interrogativos: *who*, *whom*, *whose*, *what* e *which*; já os advérbios interrogativos são: *why*, *when*, *where* e *how*.

<sup>36</sup> Nos quadros desta seção, o símbolo “%” refere-se ao percentual de ocorrências.

O fato de a seção de leitura do TOEFL ITP ter apenas dois tipos de questão resulta em um teste com pouca variedade de questões. Em contrapartida, essa baixa variedade pode ser um fator conveniente para testagens de larga escala, como é o caso dos testes TOEFL, possivelmente por ser um facilitador para a padronização do teste. A estrutura das questões desse teste é composta por apenas dois passos retóricos (comandos e alternativas, conforme mostra a Figura 8) também parece estar orientada para a otimização de aplicações do teste em larga escala.

Percebe-se também que os dois tipos de questão presentes na seção de leitura do TOEFL ITP também estão entre os tipos de questão predominantes no TESLLE de Língua Inglesa. Isso pode ser interpretado como um efeito retroativo dos testes TOEFL, sendo um indicador da influência dos testes TOEFL sobre o TESLLE de Língua Inglesa, bem como um indicador da praticidade e da recorrência desses tipos de questão no gênero.

#### 4.1.2.4 Enfoques de questão

Os enfoques das questões do TESLLE estão brevemente descritos no *site* da instituição, que diz que as questões tratam da “compreensão do texto e habilidades de leituras e [...] questões léxico-gramaticais” (UFSM, 2019). No entanto, o estudo de Marcuzzo e Radünz (2019) identificou 12 tipos de enfoques: advérbios; evidência textual; figura de linguagem; funções de partes do texto; gênero; inferência; modalidade; multimodalidade; referência; relações oracionais; tempos e vozes verbais; e tradução. Nossa análise encontrou sete destes na amostra analisada do TESLLE de Língua Inglesa, como mostra o Quadro 10.

Quadro 10 - Enfoques de questão do TESLLE.

(continua)

TESLLE de Língua Inglesa			
Enfoque	Nº	%	Excerto de questão
Evidência textual	07	43.75%	“O tema do estudo reportado no texto é” (TESLLE Q-01)

Quadro 10 - Enfoques de questão do TESLLE.

(continuação)

Referência	04	25%	“No segmento “It is <u>the last of these</u> which is challenged here” (§1), o segmento destacado se refere a” (TESLLE Q-03)
Gênero	01	6.25%	“O texto constitui um exemplar de um(a):” (TESLLE Q-14)
Funções de partes do texto	01	6.25%	“A função do segmento destacado em “Checking the results on completion of data collection or of the whole study with the respondents would, <u>it is alleged</u> , meet the requirements of diachronic reliability” (§21) é” (TESLLE Q-11)
Modalidade	01	6.25%	“Assinale a alternativa cujo fragmento sublinhado indica alto grau de certeza do autor em relação ao que afirma.” (TESLLE Q-07)
Tradução	01	6.25%	“O segmento destacado em “This appears to be an unsupported assertion that women guard <u>a self-evidently valid construct of a given phenomenon</u> ” (§17) pode ser traduzido, no contexto em que é empregado, por” (TESLLE Q-10)
Relação oracional	01	6.25%	“Assinale o segmento que <u>NÃO</u> inclui uma relação lógica de <u>CONTRASTE</u> entre as ideias.” (TESLLE Q-08)

Fonte: elaborado pelo autor.

Também são descritos seis enfoques de questões no guia oficial do TOEFL ITP, sendo esses: ideia principal; informação factual; organização e lógica; relações referenciais; vocabulário em contexto; e inferência (ETS, 2014, p. 33). Além disso, uma sétima categoria foi identificada por Perez-Amurao e chamada de “tom e propósito” (PEREZ-AMURAO, 2011, p. 53). Na amostra de TOEFL ITP analisada, revisamos essas categorias e classificamos as questões em quatro enfoques: equivalência semântico-lexical; evidência textual; referência; e inferência.

A categoria de equivalência semântico-lexical foi criada para classificar questões da seção de leitura do TOEFL ITP que identificamos abordar o significado de termos ou expressões presentes do texto-base por meio de alternativas com termos ou expressões com significado aproximado das do texto-base. O nome dessa categoria diz respeito aos aspectos linguísticos que essas questões mobilizam, uma vez que o examinando precisa entender o significado e a forma desses termos ou dessas expressões para respondê-las. Essa categoria é exemplificada pela Figura 9.

Figura 9 - Questões da seção de leitura do TOEFL ITP com enfoque em equivalência semântico-lexical

43. The word "symmetrical" in line 9 is closest in meaning to  
(A) proportionally balanced  
(B) commonly seen  
(C) typical large  
(D) steadily growing
44. The word "obvious" in line 10 is closest in meaning to  
(A) discovered  
(B) apparent  
(C) understood  
(D) simplistic

Fonte: ETS, 2014, p. 135.

As categorias de evidência textual, referência e inferência foram adotadas neste trabalho conforme as classificações e definições identificadas por Marcuzzo e Radünz (2019, p. 3648):

Evidência textual – o mesmo que “identificação de informação no texto” por Cunha (2015, p. 12) – nessas questões, é principalmente esperado que o examinando identifique evidências textuais para responder o que está sendo perguntado; [...] Referência – é esperado que o examinando identifique os referentes exofóricos, anafóricos e catafóricos de pronomes específicos, grupos nominais e pronomes, tais como “it”, “we”, “you”, “this”, “the last of these”, etc; [...] Inferência – é esperado que o examinando infira e avalie o significado de frases específicas.

Os enfoques de questão da seção de leitura do TOEFL ITP e os dados quantitativos destes estão descritos no Quadro 11.

Quadro 11 - Enfoques de questão da seção de leitura do TOEFL ITP

Seção de leitura do TOEFL ITP			
Enfoque	Nº	%	Excerto de questão
Evidência textual	21	42%	“The National Republican Party <b>is mentioned</b> in line 8 as an example of a group” (TOEFL Q-02)
Equivalência semântico-lexical	18	36%	“The word “bound” in line 1 <b>is closest in meaning to</b> ” (TOEFL Q-01)
Referência	06	12%	“The word “ones” in line 16 <b>refers to</b> ” (TOEFL Q-04)
Inferência	05	10%	“It <b>can be inferred</b> from the passage that early hotelkeepers in the United States were” (TOEFL Q-06)

Fonte: elaborado pelo autor. Grifos nossos.

De acordo com seus enfoques, as questões do TESLLE e da seção de leitura do TOEFL ITP também foram classificadas com relação aos estratos da linguagem e as competências linguísticas correspondentes.

#### 4.1.2.5 Estratos e competências das questões

A classificação das questões em estratos embasou-se na perspectiva sistêmico-funcional dos planos da linguagem (conforme apresentado anteriormente na Figura 2). Já a análise das competências exploradas pelas questões foi realizada com base em Marcuzzo e Radünz (2017; 2019).

Verificamos que as questões da amostra do TESLLE analisado remetem a quatro estratos da linguagem (Semântica e Pragmática, Discurso, Gênero e Léxico-gramática),

que correspondem a três competências linguísticas (pragmática, discursiva e linguística). Em contrapartida, todas as questões da seção de leitura do TOEFL ITP operam no estrato da Semântica e Pragmática, explorando, portanto, apenas a competência pragmática. O Quadro 12 mostra os estratos e competências envolvidas nas questões do TESLLE.

Quadro 12 - Estratos e competências envolvidas nas questões do TESLLE

TESLLE de Língua Inglesa					
Estrato	Nº	%	Competência	Nº	%
Semântica e Pragmática	13	81.25%	Pragmática	13	81.25%
Discurso	01	6.25%	Discursiva	02	12.5%
Gênero	01	6.25%			
Léxico-gramática	01	6.25%	Linguística	01	6.25%

Fonte: elaborado pelo autor.

Percebe-se que a maior variedade de enfoques de questões e de estratos da linguagem e competências mobilizadas pelo TESLLE possibilita e requer que seu texto-base, embora significativamente mais extenso, seja explorado de maneira mais completa e profunda em relação aos textos-base do TOEFL ITP. Mesmo que a seção de leitura do TOEFL ITP tenha um maior número de textos-base e de questões, os enfoques de questões e os estratos da linguagem e as competências mobilizadas por estas são menos variados que os do TESLLE. Tais fatores também colaboram para que as questões sejam repetitivas em relação ao que elas demandam, uma vez que cada um dos cinco textos-base é seguido de aproximadamente 10 questões, que se enquadram em apenas quatro dos enfoques identificados.

As questões também foram analisadas com vistas a verificar os índices de referência explícita a fragmentos do texto-base a que correspondem, como indicação de palavras, linhas, parágrafos ou passagens com informações diretamente relacionadas ao que a questão demanda. Os resultados dessa análise estão apresentados no Quadro 13.

Quadro 13 - Questões com referência explícita a fragmentos do texto-base

Questões com referência explícita a trechos do texto-base	Nº	%	Excerto de questão
TESLLE de Língua Inglesa	14	87.5%	“O tema do estudo reportado no texto é” (TESLLE Q-01)
Seção de leitura do TOEFL ITP	40	80%	“What is one of the fundamental principles of direct carving?” (TOEFL Q-09)

Fonte: elaborado pelo autor.

A partir dos resultados dessa análise foi possível perceber que grande parte das questões fazem referência explícita a trechos do texto-base a que correspondem. Questões que fazem referências ao texto-base como um todo, como, por exemplo, as questões TESLLE Q-14 “O texto constitui um exemplar de um(a) [...]” e TOEFL Q-32 “According to the passage, there is [...]” foram classificadas como questões sem referência explícita a trechos do texto-base.

Percebe-se uma sutil diferença entre os percentuais dessas ocorrências no TESLLE de Língua Inglesa e na seção de leitura do TOEFL ITP. As questões da seção de leitura do TOEFL ITP fazem ligeiramente menos referências explícitas a passagens específicas de seus textos-base correspondentes do que as questões do TESLLE de Língua Inglesa.

A partir dessas referências explícitas encontradas, constatamos que as questões de ambos os testes parecem seguir a mesma ordem ou sequência das informações apresentadas nos textos-base correspondentes. Esse resultado parece ser um recurso para facilitar a leitura dos testes, uma vez que há uma tentativa de se estabelecer uma trajetória de leitura linear paralela entre questões em relação ao textos-base que as antecedem. Há também a possibilidade de o examinando optar por ler as questões antes do seu respectivo texto-base, nesse caso, as questões irão orientar a trajetória de leitura dos textos-base de forma linear.

Além disso, em ambos os testes, as questões que não apresentam uma referência explícita a fragmentos de seus respectivos textos-base são questões com enfoque em gênero, evidência textual, e inferência.

#### 4.2 O TESLLE E O TOEFL ITP NA PRÁTICA SOCIAL

O Quadro 14 mostra o levantamento dos dados contextuais acerca dos objetivos, dos públicos-alvo, das condições de aplicação e da pontuação mínima exigida no TESLLE e no TOEFL ITP, os quais são informações relevantes para um melhor entendimento da prática social em sua totalidade.

Quadro 14 - Aspectos contextuais dos testes

(continua)

Aspectos Contextuais	TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
<b>Objetivos</b>	“[...] aferir a habilidade de leitura em língua estrangeira” (UFSM, 2019)	“A seção de leitura é projetada para aferir sua habilidade de entender passagens curtas escritas em inglês” <sup>37</sup> (ETS, 2014, p. 33)
<b>Públicos-alvo</b>	Estudantes com vínculo ativo com a UFSM. “alunos de pós-graduação, alunos de graduação que sejam prováveis formandos [...] e servidores da UFSM (docentes e técnicos administrativos em educação)” (UFSM, 2019)	Universidades; programas de internacionalização educacional; cursos livres de línguas; alunos que queiram monitorar sua proficiência no idioma e/ou participar de programas de intercâmbio acadêmico. (ETS, 2014, p. 01)
<b>Condições da aplicação</b>	Instruções orais em língua materna	Instruções orais em língua materna

<sup>37</sup> Em inglês: *The Reading Comprehension section is designed to measure your ability to understand short passages written in English.*

Quadro 14 - Aspectos contextuais dos testes

(continuação)

Aspectos Contextuais	TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
<b>Condições da aplicação</b>	Teste impresso em papel, folha de tipo A4 (21 cm de largura e 29,7 cm de altura)	Teste impresso em papel de aproximadamente meia folha A4 (cerca de 10 cm de largura e 15 cm de altura)
	São permitidas anotações no teste	Não são permitidas anotações no teste
	Aplicação semestral apenas na UFSM, sem flexibilidade de datas e horários	Aplicação livre com data, horário e local flexíveis, a depender da instituição aplicadora
	Não é possível a realização individual	Possibilidade de realização individual a depender da instituição aplicadora
	120 minutos para a realização	55 minutos para a realização da seção
	Inscrição: R\$ 53,00	Valor do teste: R\$ 520,00 <sup>38</sup>
<b>Pontuação exigida<sup>39</sup></b>	62,5%	34%

Fonte: elaborado pelo autor.

Como mostra o Quadro 14, os testes têm aspectos convergentes e divergentes, que, por repetidas vezes, se sobrepõem. Essa sobreposição pode ser percebida, por exemplo, ao comparar os objetivos desses testes. O TESLLE parece ser mais limitado, ao estabelecer (de maneira explícita) seu objetivo, e versa resumidamente apenas sobre o que pode ser interpretado como seu objetivo geral, sem detalhar de que forma(s) a leitura em língua estrangeira é aferida. Por outro lado, o objetivo da seção de leitura do TOEFL ITP, embora também seja breve, traz mais informações, indicando características

<sup>38</sup> Fonte: <https://mastertest.com.br/produto/teste-toefl-itp/>. Acesso em: 14 de jan. 2021.

<sup>39</sup> Percentual calculado com base no número mínimo de acertos exigido para a aprovação no teste e/ou consequente possibilidade de comprovação de suficiência/proficiência em leitura de EAP de acordo com os critérios da UFSM.

dos textos-base, como a multiplicidade (indicada pelo uso de plural), a extensão (passagens curtas), e o tipo de texto presente no teste (passagens escritas em língua inglesa).

Ao comparar os públicos-alvo dos testes, percebe-se o contraste entre o alcance e a dimensão destes. O TESLLE possui um público-alvo significativamente menor e menos diversificado em relação ao TOEFL ITP.

O TESLLE é realizado semestralmente por cerca de 1200 alunos da UFSM, dos quais aproximadamente 700 (o que corresponde a 58,3%) optam pelo teste de língua inglesa (MARCUIZZO, 2016). Além disso, a aplicação do TESLLE utiliza recursos públicos da UFSM. Muito embora o TESLLE não receba verbas da instituição e seja custeado pelos pagamentos dos inscritos, esse teste é aplicado nas dependências da UFSM e envolve servidores e docentes que participam dos vários processos envolvidos na testagem, como, por exemplo, a elaboração, a aplicação e a correção dos testes, além do atendimento ao público, inscrições e divulgação de resultados. Dessa forma, por utilizar, mesmo que indiretamente, recursos da UFSM, essa testagem tem um custo mais acessível que o TOEFL ITP<sup>40</sup>, além de ser projetada especificamente para atender a comunidade da própria instituição.

Por sua vez, por se tratar de uma testagem de nível internacional, o TOEFL ITP precisa dar conta de um grande número de examinandos (segundo o *site* oficial da ETS, mais de 800 mil testes são administrados anualmente). Além disso, os públicos-alvo desse teste estão inseridos em diversos contextos institucionais, como indicado pelo Quadro 14. Esses fatores, somados à necessidade da ETS de terceirizar as aplicações do teste (envolvendo a Mastertest e diversas outras instituições públicas e privadas), estão muito provavelmente relacionados ao elevado custo do teste, em comparação com o custo do TESLLE.

Os aspectos convergentes dizem respeito às instruções de ambos os testes serem realizadas em língua materna, ou seja, os aplicadores dos testes, no Brasil, fazem uso da língua portuguesa para orientar os examinandos sobre a realização e as regras dos testes. Outro ponto em comum entre esses testes é o fato de ambos serem realizados

---

<sup>40</sup> Cabe apontar que a UFSM fornece um desconto de 20% no valor do teste para os alunos da instituição, conforme notícia publicada em: <https://www.ufsm.br/2021/02/18/aplicacao-de-teste-toefl-itp-com-desconto-para-alunos-da-ufsm/>. Acesso em: 8 de mar. 2021.

fisicamente em formato impresso preto e branco em papel<sup>41</sup>. Entretanto, as dimensões da impressão são diferentes, o TOEFL ITP é impresso em formato livreto, com dimensões de aproximadamente meia folha A4 (cerca de 10 cm de largura e 15 cm de altura), ao passo que o TESLLE é impresso em folhas A4 (21 cm de largura e 29,7 cm de altura)<sup>42</sup>. Dessa forma, é possível argumentar que o TESLLE possui características físicas que podem vir a ser elementos facilitadores da leitura de ESP, tendo em vista que a impressão convencional do TESLLE é quase duas vezes maior em relação ao TOEFL ITP, além de serem permitidas anotações nas páginas do teste. No entanto, as formas de aplicação do TOEFL ITP são mais complacentes que as do TESLLE, pois existe a possibilidade de se realizar o TOEFL ITP individualmente, com flexibilidade de data, horário e local, a depender da instituição aplicadora, diferentemente do TESLLE, que é aplicado (até o presente) apenas no campus sede e no campus de Frederico Westphalen da UFSM, semestralmente e de forma coletiva.

A seção de leitura do TOEFL ITP tem duração de 55 minutos e possui cinco textos-base com extensão de entre 300 e 350 palavras e 50 questões de múltipla-escolha com quatro alternativas. Dessa forma, há, em média, 66 segundos para que se responda cada uma das 50 questões. Em contrapartida, o TESLLE tem duração de 120 minutos e apenas um texto-base com extensão de 28 parágrafos e 16 questões de múltipla-escolha com cinco alternativas, o que resulta em uma média de sete minutos e 30 segundos para cada questão. Tais fatores viabilizam que os examinandos possam fazer uma leitura mais cuidadosa e, assim, possam responder questões mais complexas sobre o texto-base do TESLLE, mesmo que esse seja mais extenso que os da seção de leitura do TOEFL ITP.

O tempo mais limitado para que se responda às questões da seção de leitura do TOEFL ITP, se interpretado juntamente com o número de tipos e enfoques destas, demonstra que as questões desse teste promovem uma exploração mais superficial dos textos-base em relação ao TESLLE. Observamos que questões do TESLLE são significativamente menores em número, porém mais variadas (de acordo com nossas análises dos tipos, enfoques, estratos e competências). Além disso, são disponibilizados,

---

<sup>41</sup> Recentemente, por conta da pandemia do Coronavírus (COVID-19), o TESLLE e o TOEFL ITP passaram a ser realizados também em formato digital.

<sup>42</sup> Essas informações foram obtidas a partir de um estudo sobre multimodalidade, realizado pelo autor, e que está em vias de ser publicado, que investigou o sistema composicional da amostra do TESLLE de Língua Inglesa.

em média, sete minutos e 30 segundos para que se respondam estas questões. Ao cruzar os números correspondentes ao tempo para realização e a quantidade de questões, de tipos, de enfoques e de textos-base, além de não ser permitido fazer anotações no teste, pode-se construir um argumento sólido de que o TOEFL ITP é um teste no qual habilidades como as de concentração, gerenciamento de tempo e memorização temporária de informações podem ser tão importantes e decisivas para um bom desempenho dos examinandos quanto às de compreensão escrita. Esse argumento é reforçado pelo fato de que, além de compreender o texto-base de acordo com o que cada questão demanda, os examinandos precisam fazê-lo repetidas vezes em um curto período de tempo, de maneira exaustiva e sem utilizar de recursos visuais, como, por exemplo, destacar passagens do texto-base ou fazer pequenas anotações paralelas. Esses recursos visuais, que são permitidos apenas no TESLLE, podem ser facilitadores para a memorização de informações e para a transição de leitura entre as páginas dos textos-base, das questões que os sucedem e até mesmo da folha-resposta. A proibição desses recursos parece implicar um maior esforço de memorização e para (re)encontrar informações específicas no texto-base.

Por fim, outra questão importante para o entendimento dessa prática social é a pontuação<sup>43</sup> (*scores*) esperada ou exigida para o sucesso dos examinandos. De acordo com os editais do TESLLE, para ser aprovado no TESLLE (e posteriormente comprovar suficiência em leitura de uma língua estrangeira na UFSM), um aluno deve obter uma nota igual ou superior a 6,1, ou o conceito B-, que equivale a 10 acertos nas 16 questões do TESLLE, ou seja, responder corretamente 62,5% das questões do teste. “Para cada nota, há um conceito correspondente, definido pelo Regimento Geral da Pós-graduação, da UFSM, na seção IV, art. 60” (TESLLE, 2020, p. 8). O quadro a seguir apresenta essa correspondência entre o número de acertos, a nota e os conceitos, de acordo com os critérios do TESLLE e do Regimento Geral da Pós-graduação da UFSM (2014).

---

<sup>43</sup> Em português também são utilizados os termos nota, score e conceito, a depender do teste e instituição responsável.

Quadro 15 - Scores do TESLLE

Intervalo de nota		Conceitos	Número de acertos no teste	Situação do examinando
0,00	1,00	E-	0 a 1	REPROVADO
1,10	2,00	E	2 a 3	
2,10	3,00	D-	4	
3,10	4,00	D	5 a 6	
4,10	5,00	C-	7 a 8	
5,10	6,00	C	9	APROVADO
6,10	7,00	B-	10 a 11	
7,10	8,00	B	12	
8,10	9,00	A-	13 a 14	
9,10	10,00	A	15 a 16	

Fonte: Adaptado de TESLLE (2020, p. 8).

A pontuação da seção de leitura do TOEFL ITP, por sua vez, é de entre 31 e 67 pontos. Isso significa que, ainda que um examinando responda todas as 50 questões incorretamente, sua pontuação será de 31. Dessa forma, não é possível uma nota zero, por exemplo, além de não haver aprovação ou reprovação nos testes TOEFL, ou uma média estipulada para isso. No entanto, a pontuação nos testes TOEFL segue e se enquadra no QCER, que é amplamente usado como um indicador de proficiência linguística. Assim, a pontuação na seção de leitura do TOEFL ITP é classificada da seguinte forma:

Quadro 16 - Descrições de proficiência de acordo com o nível QCER e a pontuação na seção de leitura do TOEFL ITP

(continua)

Pontuação na seção de leitura do TOEFL ITP	Nível QCER	Descrição da proficiência
63-67	C1	Examinandos neste nível são normalmente capazes de: <ul style="list-style-type: none"> <li>acompanhar discursos no nível da ideia para entender informações detalhadas e ideias principais, afirmadas tanto explicitamente quanto implicitamente, mesmo quando: <ul style="list-style-type: none"> <li>textos contêm acúmulo de vocabulário acadêmico utilizado com baixa frequência</li> <li>comparações e contrastes, relações causais, ilustrações, etc. que não são afirmadas explicitamente ou indicadas por marcadores discursivos</li> </ul> </li> </ul>

Quadro 16 - Descrições de proficiência de acordo com o nível QCER e a pontuação na seção de leitura do TOEFL ITP

(continuação)

Pontuação na seção de leitura do TOEFL ITP	Nível QCER	Descrição da proficiência
63-67	C1	- textos são sobre tópicos abstratos, como composição musical e animação de computador
56-62	B2	Examinandos neste nível são normalmente capazes de: <ul style="list-style-type: none"> <li>• Processar informações entre textos acadêmicos típicos para entender informações detalhadas e ideias principais, afirmadas tanto explicitamente quanto implicitamente, mesmo quando os textos: <ul style="list-style-type: none"> <li>- contém vocabulário acadêmico utilizado com alta frequência e marcadores típicos de discurso acadêmico</li> <li>- são sobre tópicos concretos que discutem as ciências físicas e sociais (e.g., formação glacial, relevo lunar, teorias de desenvolvimento infantil)</li> </ul> </li> </ul>
48-55	B1	Examinandos neste nível são normalmente capazes de: <ul style="list-style-type: none"> <li>• entender descrições de processos relativamente simples e narração em textos acadêmicos bem marcados</li> <li>• entender vocabulário de uso frequente e reconhecer informações parafraseadas</li> <li>• acompanhar comparações e contrastes no nível da frase e entender o significado expressado pelas conjunções mais comuns como “and”, “or” e “but”</li> <li>• fazer conexões de significado entre algumas orações simples que contém vocabulário de uso frequente</li> </ul>
31-47	A2	Examinandos neste nível são às vezes capazes de: <ul style="list-style-type: none"> <li>• entender a ideia geral de algumas frases que usam vocabulário simples do cotidiano</li> <li>• entender a ideia principal de alguns textos nos quais a ideia é reforçada pela repetição de vocabulário importante ao longo de várias frases</li> <li>• acompanhar referências de frases simples (e.g., “it”, “they”) para determinar o referente gramatical de um pronome</li> <li>• localizar informação solicitada em algumas frases se apontado diretamente para a parte da passagem que contém a informação (e.g., “in line x,” “in paragraph y”)</li> </ul>

Fonte: Adaptado de TOEFF ITP *Test Score Descriptors* (2014b, p. 3). Tradução do autor.

Essa pontuação corresponde diretamente ao número de acertos na seção de leitura do TOEFL ITP, como mostra a Figura 10.

Figura 10 - Equivalência do número de acertos em pontos na seção de leitura do TOEFL ITP

Tabela de Conversão			
Total de acertos	Seção 1 - Listening (Convertida em pontos)	Seção 2 - Grammar (Convertida em pontos)	Seção 3 - Reading (Convertida em pontos)
48-50	64-68		65-67
45-47	61-63		61-63
42-44	58-60		58-60
39-41	55-57		56-58
36-38	53-55	63-68	54-56
33-35	51-53	59-61	53-54
30-32	50-51	56-58	51-52
27-29	48-49	53-55	49-51
24-26	47-48	50-52	48-49
21-23	45-46	48-49	45-47
18-20	43-44	45-47	43-45
15-17	41-42	42-44	39-42
12-14	37-40	39-42	32-37
9-11	33-35	34-38	31
6-8	31-32	31	31
0-5	31	31	31

Fonte: Idioma sem Fronteiras da Universidade Federal de Uberlândia (2016).

Instituições que fazem uso ou aceitam os resultados dos testes TOEFL como comprovação para determinados fins estipulam pontuações mínimas a serem obtidas de acordo com seus próprios critérios. Na UFSM, para requisitar a comprovação de suficiência em leitura de língua inglesa através do resultado do TOEFL ITP, os alunos são orientados pelo site do TESLLE/UFSM, na aba de perguntas frequentes, da seguinte forma:

**Posso solicitar a revalidação da nota do TOELF como teste de suficiência?**

Sim. Para obter a equivalência, o requerente terá que ter nota igual ou superior a 41 pontos na seção “Reading comprehension” do TOELF [ITP]. Deve-se redigir um requerimento assinado com o cartão de resultado do TOELF em anexo e encaminhar para o programa de pós-graduação do requerente<sup>44</sup>.

<sup>44</sup> Fonte: <https://www.ufsm.br/pro-reitorias/prpgp/teslle/perguntas-frequentes/>. Acesso em: 7 fev. 2021.

Assim, para obter a comprovação de suficiência na UFSM através do TOEFL ITP, o examinando precisa acertar cerca de 17 questões na seção de leitura do teste, o que corresponde a 34% das questões da seção.

Dessa forma, ao comparar o percentual mínimo de acertos no TESLLE e na seção de leitura do TOEFL ITP exigido pela UFSM para a comprovação de suficiência, percebe-se a disparidade entre o índice de 62,5% de acertos necessários no TESLLE e os 34% necessários na seção de leitura do TOEFL ITP. Em termos quantitativos, isso significa que de acordo com os critérios da UFSM, é relativamente mais “fácil” de se obter a comprovação por meio da realização do TOEFL ITP do que por meio da realização TESLLE. No entanto, assim como todos os resultados desta pesquisa, essa diferença entre os percentuais mínimos de acerto exigidos em cada teste não pode ser interpretada isoladamente.

Para que se tenha um entendimento tão completo quanto possível dos testes e da prática de testagem aqui investigados, a seção a seguir busca sistematizar as numerosas diferenças textuais e contextuais entre esses testes detalhadas e discutidas neste trabalho.

#### 4.3 SISTEMATIZAÇÃO E INTERPRETAÇÃO DOS RESULTADOS

O Quadro 17 foi elaborado de forma a resumir os resultados das análises do texto e do contexto traçando um paralelo entre os aspectos dos testes analisados pelo presente estudo.

Quadro 17 - Síntese dos resultados

(continua)

Aspectos investigados		TESLLE de Língua Inglesa	Seção de leitura do TOEFL ITP
Texto	Teste		
	Movimentos retóricos	4	7
	Quantidade	1	5
Textos-base	Extensão	Cerca de 5 mil palavras	Entre 300 e 350 palavras

Quadro 17 - Síntese dos resultados

		(continuação)		
<b>Texto</b>		Quantidade	16	50
	<b>Questões</b>	Estrutura	Entre 2 e 5 passos retóricos	2 passos retóricos
		Tipos	5	2
		Enfoques	7	4
		Estratos da linguagem	4	1
		Competências linguísticas	3	1
		Referência explícita a trechos do texto-base	87.5%	80%
		<b>Objetivos do teste</b>	“[...] aferir a habilidade de leitura em língua estrangeira” (UFSM, 2019)	“[...] aferir sua habilidade de entender passagens curtas escritas em inglês” (ETS, 2014, p. 33)
<b>Público-alvo</b>	Local (comunidade acadêmica da UFSM)	Internacional (alunos e instituições diversas)		
<b>Contexto</b>	<b>Condições da aplicação</b>	Instruções orais do aplicador	Em língua materna	Em língua materna
		Formato físico do teste	Folha de tipo A4	Aproximadamente meia folha de tipo A4
		Anotações no teste	São permitidas	Não são permitidas
		Aplicação	Coletiva; semestral	Coletiva ou individual; data e horário marcados
		Tempo para realização (média por questão)	120 minutos (média de 7 minutos e 30 segundos)	55 minutos (média de 66 segundos)
		Custo para o examinando	R\$ 53,00	R\$ 520,00
		<b>Média percentual de acertos exigida pela UFSM</b>	62,5% (10 das 16 questões)	34% (cerca de 17 das 50 questões)

Fonte: elaborado pelo autor.

Assim, ao compararmos os aspectos dos testes na prática social, podemos apontar os fatores que influenciam a confiabilidade e a validade desses testes, e, a partir disso, discutir os possíveis efeitos retroativos e mapear as concepções de linguagem, proficiência em leitura de EAP e letramento que subjazem a esses testes.

No que diz respeito aos textos dos testes, o TESLLE de Língua Inglesa apresenta uma maior variedade de questões em todos os aspectos (tipos, enfoques, estratos da linguagem e competências linguísticas envolvidas) em relação à seção de leitura do TOEFL ITP, que, por sua vez, tem um maior número de textos-base e de questões. Contudo, os textos-base da seção de leitura do TOEFL ITP são significativamente mais curtos que o texto-base do TESLLE de Língua Inglesa, que tem cerca de três vezes mais palavras que os cinco textos-base do TOEFL ITP juntos.

As questões do TESLLE também apresentam um percentual ligeiramente maior de ocorrências de referência explícita a trechos do texto-base que as do TOEFL ITP, que pode ser interpretado como um facilitador para o respondimento das questões. Ainda assim, as questões do TESLLE, diferentemente das da seção de leitura do TOEFL ITP, contemplam outros estratos da linguagem, além do da semântica e pragmática, incluindo estratos mais abstratos como os do gênero e do discurso, bem como o estrato da léxico-gramática.

Além disso, as questões do TESLLE de Língua Inglesa são estruturalmente mais complexas que as da seção de leitura do TOEFL ITP. As questões do TESLLE podem apresentar até cinco passos retóricos em sua estrutura (contextualização; primeiro comando; dados; segundo comando; e alternativas de resposta, conforme Figura 8), ao passo que as da seção de leitura do TOEFL ITP apresentam sempre apenas dois passos retóricos (comando e alternativas de resposta). Essa maior variedade e complexidade na formulação das questões do TESLLE de Língua Inglesa reforça nossa interpretação de que as questões desse teste são potencialmente mais difíceis do que as da seção de leitura do TOEFL ITP.

Ademais, a seção de leitura do TOEFL ITP apresenta um maior número de movimentos retóricos que o TESLLE quando analisados como gêneros discursivos. A partir desse resultado, podemos considerar que a seção de leitura do TOEFL ITP é um

gênero mais complexo que o TESLLE de Língua Inglesa. No entanto, interpretamos essa maior complexidade do TOEFL ITP como um indicador da alta padronização que busca facilitar a aplicação do teste, uma vez que os movimentos retóricos presentes apenas na seção de leitura do TOEFL ITP (instruções iniciais; exemplos de texto-base e questões; e instruções finais) são elementos facilitadores para que os examinandos entendam o funcionamento do teste.

Quanto aos aspectos contextuais, percebem-se diferenças significativas entre os objetivos dos testes. Ao analisar com atenção o objetivo da seção de leitura do TOEFL ITP, que é “[...] aferir sua habilidade de entender passagens curtas escritas em inglês” (ETS, 2014, p. 33), é seguro assumir que o termo entender diz respeito a compreensão dos aspectos da semântica e pragmática do texto-base, uma vez que todas as questões desse teste envolvem esse estrato da linguagem. Nesse sentido, o objetivo do TESLLE de Língua Inglesa de “[...] aferir a habilidade de leitura em língua estrangeira” (UFSM, 2019) parece ser mais coerente (ser mais válido) para um teste de leitura, no sentido em que a habilidade de leitura é mais complexa que a habilidade de entender. Essa diferença não é apenas semântica entre os termos ler e entender. A leitura, como já mencionado, é um processo interativo entre autor e leitor no qual esse busca obter significado de um texto (NUTTALL, 1996, p. 4), e, portanto, ler é um processo sociocomunicativo que vai além do entendimento de um texto em língua estrangeira.

Ainda comparando os objetivos desses testes, aferir a “habilidade de entender passagens curtas escritas em inglês” (ETS, 2014, p. 33), objetivo da seção de leitura do TOEFL ITP, parece ser menos válido para o contexto da prática social da testagem de proficiência em leitura de EAP na UFSM do que o objetivo do TESLLE, que é mais abrangente ao buscar aferir a “leitura em língua estrangeira” (UFSM, 2019). Essa menor validade da seção de leitura do TOEFL ITP para esse contexto é reforçada por, pelo menos, outros dois fatores: i) a já discutida não autenticidade dos textos-base desse teste; e ii) o público-alvo mais amplo e, portanto, mais variado, fator que demanda que o teste seja mais abrangente e atenda às necessidades de comunidades bastante distintas, prejudicando também a confiabilidade do teste (grau de precisão da aferição do que se deve aferir), uma vez que “passagens curtas” são textos abrangentes o suficiente para o público-alvo desse teste, mas consideravelmente vagos para o contexto acadêmico.

Há ainda, de acordo com Wielewicki (1996, p. 61), fatores aleatórios que podem prejudicar indiretamente a pontuação e, portanto, a confiabilidade (grau de precisão da aferição) de um teste, como o nível de tensão dos examinandos. Esse fator não pôde ser mensurado objetivamente no presente estudo, mas podemos apontar alguns aspectos dos testes que podem contribuir para o aumento da tensão dos examinandos e, conseqüentemente, uma redução na confiabilidade do teste. No caso do TESLLE, podemos apontar o percentual de acertos mínimo exigido pela UFSM, que é mais de duas vezes maior que o percentual exigido na seção de leitura do TOEFL ITP, e a aplicação necessariamente coletiva e semestral do TESLLE, que pode, porventura, ser um fator inconveniente para um candidato que tenha dificuldade em se concentrar com a presença de dezenas de outros examinandos na mesma sala, e que, caso reprove no teste, precisará esperar até o semestre seguinte para realizar o teste novamente. Esse fator pode ser evitado na realização do TOEFL ITP, que pode ocorrer de forma individual e com um intervalo menor, caso seja necessário repetir o teste.

Todavia, a seção de leitura do TOEFL ITP também apresenta, pelo menos, três fatores que podem prejudicar a confiabilidade do teste por serem ao menos potencialmente desfavoráveis para os examinandos, a saber: i) a proibição de anotações no teste; ii) o tempo médio mais limitado para que se responda às questões (cerca de oito vezes menor que o tempo disponibilizado pelo TESLLE); e iii) o elevado custo de inscrição (quase dez vezes maior que o valor cobrado para a realização do TESLLE).

Por fim, foram observados dois fatores comuns a ambos os testes que podem contribuir para um maior índice de confiabilidade destes: i) as instruções orais fornecidas pelo aplicador dos testes são feitas em língua materna, o que facilita a comunicação entre aplicador e examinando e, conseqüentemente, reduz as chances de que haja dúvidas a respeito da realização do teste; e ii) os elevados índices de questões que fazem referência explícita a trechos do texto-base, indicando a(s) linha(s) que abordam, ainda que parcialmente, as informações necessárias para que se responda à questão, facilitando assim que os examinandos localizem informações específicas nos textos-base.

Desse modo, parece seguro afirmar que o TESLLE de Língua Inglesa e o a seção de leitura do TOEFL ITP têm concepções distintas de linguagem, proficiência em leitura

de EAP e (multi)letramento. O TESLLE de Língua Inglesa, cuja concepção é de “linguagem como gênero” (MARCUIZZO; RADÜNZ, p. 3652), valoriza a leitura de textos autênticos, que são tipicamente lidos pela comunidade acadêmica, além de apresentar questões com formulações mais complexas (com relação a seus tipos e enfoques) e que concernem a linguagem de forma mais abrangente (abrangendo quatro dos seis estratos da linguagem). Ao passo que, para a seção de leitura do TOEFL ITP, considerando os textos não autênticos, somados a questões estruturalmente simples e que tratam apenas do extrato da semântica e da pragmática, pode-se concluir que as habilidades de memorização e de gerenciamento de tempo são centrais para a concepção de linguagem e de proficiência em leitura de EAP pressuposta pelo teste.

Dessa forma, também pode-se concluir que o letramento visual é um aspecto testado, em certa medida, pelo TESLLE de Língua Inglesa, já que esse teste contém um texto-base autêntico, inclusive com relação a seus aspectos visuais (*layout*). A leitura dos significados visuais que remetem ao gênero do texto-base, por exemplo, pode ser decisiva para que se responda à questão TESLLE Q-14: “O texto constitui um exemplar de um(a):”, cuja resposta correta é a alternativa A: “Artigo de revisão”. Assim, parece seguro afirmar que a concepção de multiletramentos subjaz, ainda que superficialmente, ao TESLLE de Língua Inglesa. No entanto, o mesmo não pode ser afirmado sobre a seção de leitura do TOEFL ITP, cujos textos-base e questões não apresentam ou abordam significados não verbais. Isso demonstra que a concepção de letramento subjacente à seção de leitura do TOEFL ITP desconsidera os múltiplos significados não verbais que constituem a concepção de multiletramentos.

Em resumo, ser proficiente em leitura de EAP para o TESLLE é ler textos autênticos, ainda que complexos e extensos, e compreender não somente os extratos da léxico-gramática e da semântica e pragmática do texto, mas também aspectos do gênero e seus propósitos sociocomunicativos, e do(s) discurso(s) que subjaz(em) ao gênero em questão. Ao passo que, para a seção de leitura do TOEFL ITP, proficiência em leitura de língua inglesa significa entender com imediatez e objetividade os aspectos semânticos e pragmáticos de textos breves e descontextualizados.

Dessa forma, podemos estimar os efeitos retroativos, ou seja, as influências potenciais desses testes e dessa prática de testagem tanto no contexto educacional

quanto social. É necessário notar que pesquisas posteriores, de caráter etnográfico, por exemplo, com os participantes da prática social são necessárias para verificar essas estimativas. Essa discussão relaciona os resultados desta pesquisa com as cinco dimensões do conceito de efeito retroativo levantadas pelo estudo de Scaramucci (2004).

1. Especificidade: geral e específico – ambos os testes parecem ter efeito retroativo geral e específico, uma vez que ambos podem induzir os examinandos e também professores que ministram aulas preparatórias para estes testes a estudar tanto o conteúdo dos testes, como estratégias de leitura, por exemplo, quanto os aspectos técnicos dos testes em si, como suas estruturas e regras de realização, por exemplo, que são fatores de interesse dos examinandos e tipicamente abordados por cursos e materiais preparatórios.
2. Intensidade: forte – na própria UFSM existem cursos voltados exclusivamente para preparar examinandos para o TESLLE de Língua Inglesa e o TOEFL ITP, oferecidos pelo IsF e LinC, respectivamente. Além disso, a aprovação nos testes de proficiência ou suficiência em leitura de EAP também é central para aulas de cursos livres de língua inglesa e de professores particulares. Portanto, o efeito retroativo de intensidade de ambos os testes e da prática social é forte.
3. Extensão: curto e longo – essa dimensão do efeito retroativo da prática social aqui investigada pode ser considerada curta, pois os efeitos de ambos os testes não necessariamente permanentes, uma vez que o objetivo do público da UFSM que realiza esses testes é, em geral, apenas obter a comprovação de proficiência ou suficiência para cumprir essa exigência institucional. No mesmo sentido, o efeito retroativo pode ser considerado longo uma vez que as concepções de linguagem desses testes podem exercer influência na concepção de linguagem dos examinandos que os realizam. No entanto, essas especulações devem ser revisadas em estudos posteriores.
4. Intencionalidade: intencional e não intencional – há consequências sociais intencionais e não intencionais de ambos os testes, que, novamente, precisam ser exploradas de maneira mais aprofundada por outras pesquisas. De qualquer forma, um possível exemplo que podemos apontar na condição de participantes dessa prática social é o amplo e conhecido mercado de cursos e materiais

preparatórios oficiais e não oficiais que se criou a partir dos testes TOEFL. Com o TESLLE, no entanto, isso não ocorre por se tratar de um teste local e não haver material ou curso preparatório oficial para esse teste. No entanto, sabe-se da influência do TESLLE de Língua Inglesa em projetos de laboratórios de ensino como o LinC que oferece semestralmente um curso preparatório que inclui um livro didático para a comunidade acadêmica da UFSM. Esse curso e material são elaborados, organizados e ministrados pelos membros do projeto, que são alunos de graduação e Pós-graduação em Letras e professores do DLEM. Assim, consolida-se uma prática social (local) de ensino centrada no TESLLE de Língua Inglesa. Também é importante destacar que são necessárias investigações a respeito da intencionalidade ou não intencionalidade dos efeitos retroativos gerados pela significativa diferença entre as médias percentuais de acertos exigidas para cada um dos testes. Embora não haja descritores oficiais dos índices de suficiência do TESLLE, a pontuação mínima exigida pela UFSM na seção de leitura do TOEFL ITP (41) corresponde ao nível A2 do QCER. Os descritores apontam que os examinandos enquadrados nesse nível são, por exemplo, “às vezes capazes de entender a ideia geral de algumas frases que usam vocabulário simples do cotidiano” (ETS, 2014b, p. 3), o que parece ser um nível de proficiência em leitura de EAP significativamente inferior ao nível de examinandos que obtêm aprovação no TESLLE de Língua Inglesa, além de ser um nível que dificilmente seria considerado “suficiente” para acadêmicos de pós-graduação, que frequentemente precisam ler textos científicos complexos em língua inglesa. De todo modo, esse aspecto da prática de testagem pode ser discutido em estudos que incluam entrevistas com participantes dessa prática que tenham algum nível de responsabilidade pela tomada de decisões como a média percentual de acertos a ser exigida em cada teste.

5. Valor: positivo e negativo – para os examinandos, os fatores positivos que se destacam com relação ao TESLLE são: i) o baixo valor de inscrição; ii) a possibilidade de fazer anotações no teste; iii) o maior tempo médio para responder às questões; e iv) as questões serem escritas em língua portuguesa. Esses fatores são negativos quando se trata da seção de leitura do TOEFL ITP (ver Quadro 17).

Com relação ao TOEFL ITP, cabe mencionar: i) a flexibilidade de data e horário para a realização do teste; ii) os exemplos de texto-base e questões (didatização); iii) a menor variedade de tipos, enfoques, estratos e competências envolvidos pelas questões; e iv) a menor média percentual de acertos exigida pela UFSM. Esses fatores são negativos quando se trata do TESLLE de Língua Inglesa (ver Quadro 17). Ademais, os fatores positivos comuns a ambos os testes são: i) o elevado número de questões que fazem referência explícita a trechos do texto-base; e ii) as instruções orais do aplicador serem feitas em língua materna.

O próximo capítulo apresenta as conclusões finais e implicações e limitações práticas e teóricas desta pesquisa.

## CAPÍTULO 5 - CONSIDERAÇÕES FINAIS

O presente estudo buscou investigar a prática de testagem de proficiência em leitura de EAP no âmbito da UFSM e, para isso, buscamos descrever as principais características dos textos e contextos dos testes TESLLE e da seção de leitura do TOEFL ITP. Com relação à análise do contexto, foram identificados e descritos os objetivos comunicativos, os públicos-alvo, as condições de aplicação e as médias desses testes. No que se refere à estrutura, foram investigados os movimentos retóricos do gênero teste de leitura de EAP, as características dos textos-base (gêneros, extensão e temas abordados), os tipos e enfoques das questões, além dos estratos da linguagem e as competências linguísticas que essas questões mobilizam. Todos esses fatores exercem influência nos processos da testagem, e a identificação deles tornou possível discutir a confiabilidade, a validade e o efeito retroativo desses testes na prática social.

A importância de se investigar o contexto para as análises de texto e linguagem e o papel educacional e social de testes são frisadas por Bachman (1990, p. 279) “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”<sup>45</sup>. Nesse sentido, este estudo parte do princípio de que a(s) análises de) linguagem não é (são) meramente uma forma de (analisar) representação, mas de ação social (MILLER, 1984). Ao analisar as características linguísticas, discursivas e sociais, esta pesquisa busca contribuir para possíveis recontextualizações ou alterações do gênero teste de leitura de EAP para fins pedagógicos, por exemplo. Essas mudanças podem ser realizadas por ou a partir de estudos que cubram as limitações deste e poderão proporcionar benefícios para a comunidade acadêmica da UFSM que realiza esses testes e para todos os participantes das práticas sociais de testagem de leitura de EAP, que são comuns a diversos contextos acadêmicos internacionais.

As implicações práticas e teóricas deste estudo também visam a beneficiar e se somar aos estudos de linguagem, especialmente àqueles da área de Linguística

---

<sup>45</sup> *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.* Tradução do autor.

Aplicada, com enfoque em Análise (Crítica) de Gênero, análises de práticas sociais, testagem e em inglês para fins específicos. Esperamos que, com o amparo de pesquisas de linguagem e da ciência como um todo, testes e práticas sociais de testagem como as investigadas e discutidas neste estudo possam ser otimizadas em todos os sentidos para que, idealmente, atuem como instrumentos educacional e socialmente benéficos.

## REFERÊNCIAS

ALDERSON, J.C.; CLAPHAM, C.; WALL, D. **Language test construction and evaluation**. Cambridge, Cambridge University Press. 1995.

BAZERMAN, C. **Shaping written knowledge**: the genre and activity of the experimental article in science. Madison: University of Wisconsin Press, 1988.

BECKER, L. W. **EFL Reading patterns**: a study based on Teste de Suficiência at the Federal University of Santa Maria. 2013. 19f. Trabalho Final de graduação (Curso de Graduação em Letras Português-Inglês) Universidade Federal de Santa Maria, Santa Maria, RS, 2013.

BHATIA, V. **Analysing genre**: language use in professional settings. London: Longman, 1993.

BRAHIM, A. C. S. M. Do estruturalismo à concepção discursiva de língua: os impactos na pedagogia de leitura em língua estrangeira. **Revista Intersaberes**, v. 8, n. 16, p.108-130, 2013.

CELCE-MURCIA, M.; DÖRNYEI, Z.; THURRELL, S. Communicative competence: a pedagogically motivated model with content specifications. **Issues in Applied Linguistics**, v. 6, n. 2, p. 5-35, 1995.

CELIA, M. H. C. **Leitura em inglês**: formato alternativo para um exame de proficiência. 1983. 278 p. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre. 1983.

CHAPELLE, C.A. Validity in language assessment. **Annual Review of Applied Linguistics**, v. 19, p. 254-272. 1999.

COPE, B.; KALANTZIS, M. **Multiliteracies**: new literacies, new learning. *Pedagogies: An International Journal*, v. 4, n. 3, p. 164-195, 2009.

CUNHA, A. C. S. **The reading perspective of TESLLE**: the English Reading proficiency test of UFSM. 2015. 19f. Trabalho Final de Graduação (Curso de Graduação em Letras Inglês) Universidade Federal de Santa Maria, Santa Maria, RS, 2015.

EBEL R.L.; FRISBIE, D. A. **Essentials of Educational Measurement**. 5 ed., Englewood Cliffs, Prentice-Hall, 1991.

EDUCATIONAL TESTING SERVICE. **Official TOEFL handbook**. Princeton, 2012a.

EDUCATIONAL TESTING SERVICE. **Supervisor's manual**. Princeton, 2012b.

EDUCATIONAL TESTING SERVICE. **Official guide to the TOEFL ITP test**. Princeton, 2014a.

EDUCATIONAL TESTING SERVICE. **TOEFF ITP Test Score Descriptors**. Princeton, 2014b. Disponível em: [https://www.ets.org/s/toefl\\_itp/pdf/test\\_score\\_descriptors.pdf](https://www.ets.org/s/toefl_itp/pdf/test_score_descriptors.pdf). Acesso em: 15 fev. 2021.

EDUCATIONAL TESTING SERVICE. **Quem somos**. Disponível em: <http://mastertest.com.br/quem-somos/>. Acesso em: 16 fev. 2020a.

EDUCATIONAL TESTING SERVICE. **Reading comprehension sample questions**. Disponível em: [https://www.ets.org/toefl\\_itp/content/sample\\_questions/level1\\_section3\\_reading\\_comprehension](https://www.ets.org/toefl_itp/content/sample_questions/level1_section3_reading_comprehension). Acesso em: 15 mar. 2020b.

FAIRCLOUGH, N. **Language and power**. London: Longman, 1989.

FAIRCLOUGH, N. **Discourse and social change**. Cambridge: Polity Press, 1992.

FAIRCLOUGH, N. **Media discourse**. London: Edward Arnold, 1995.

FAIRCLOUGH, N. **Discurso e mudança social**. Trad. Isabel Magalhães (Coord.). Brasília: Editora UnB, 2001.

FAIRCLOUGH, N. **Critical discourse analysis**. 2 ed. Harlow, UK: Pearson, 2010.

FINARDI, K. R.; FRANÇA, C. O Inglês na internacionalização da produção científica brasileira: evidências da subárea de linguagem e linguística. **Revista Intersecções**, v. 19, p. 234–250, 2016.

FINARDI, K. R. English as a global language in Brazil: a local contribution. *In*: T. Gimenez, M. S. El Kadri, & L. C. S. Calvo (Eds.), **English as a lingua franca in teacher education: a Brazilian perspective**. Berlin: De Gruyter Mouton. p. 71–86, 2018.

FLOREK, C. S. **Um resumo acadêmico gráfico vale mais do que mil palavras**: análise crítica de um gênero multimodal polivalente. 2018. 255 p. Tese (Doutorado em Estudos Linguísticos) – Universidade Federal de Santa Maria, Santa Maria, 2018.

FREIRE, P. **Educação e mudança**. 12ª Edição. Paz e Terra. Rio de Janeiro, 1979.

FUZER, C. et al. **Diretrizes para elaboração e revisão de questões em processos seletivos**. Santa Maria: Programa de Pós-Graduação em Letras — PPGL/UFSM, 2017.

HALLIDAY, M. A. K. **Language as social semiotic**: the social interpretation of language and meaning. London: Edward Arnold, 1978.

HALLIDAY, M. A. K. **An introduction to functional grammar**. London: Edward Arnold, 1994.

HALLIDAY, M. A. K.; HASAN, R. **Language, context, and text: aspects of language in a social-semiotic perspective.** Oxford: Oxford University Press, 1989.

HALLIDAY, M. A. K.; MATTHIESSEN, C. **An introduction to functional grammar.** 4 ed. London: Routledge. 2004/2014.

HENDGES, G. R. **Gramática visual: significados representacionais.** 2019. 138 diapositivos, color.

HUGHES, A. **Testing for Language Teachers.** Cambridge. Cambridge University Press. 1989.

HUGHES, A. Teaching and testing. In: HUGHES, A. **Testing for language teachers.** Cambridge: Cambridge University Press, 1997. p. 1-5.

JENKINS, J. **English as a lingua franca in the international university: the politics of academic English language policy.** New York: Routledge, 2014.

KAMIJO, T. Analyzing the TOEFL ITP reading comprehension from a test preparation textbook: a case study and its classroom application. **Journal of International Institute of Language and Cultural Studies.** Ritsumeikan University, v. 21, n. 4, p. 195-206, 2009.

KAWACHI-FURLAN, C. J.; AMORIM, G. B.; FINARDI, K. R. The interface between TOEFL ITP and internationalization and language assessment in Brazil. **Studies in English language teaching,** v. 5, n. 2, p. 213-230, 2017.

KLEIMAN, A. **Os significados do letramento: uma nova perspectiva sobre a prática social da escrita.** Campinas: Mercado da Letras, 1995.

KRESS, G. Visual and verbal modes of representation in electronically mediated communication: the potential of new forms of text. In: SNYDER, I. **Page to screen: taking literacy into the electronic era.** London: Routledge, 1997. p. 53-59.

KRESS, G.; van LEEUWEN, T. **Reading images: the grammar of visual design.** 2. ed. London: Routledge, 2006.

LITTLE, D; DEVITT, S; SINGLETON, D. **Learning foreign languages from authentic texts: theory and practice.** Dublin: Authentik, 1989.

LITTLEWOOD, W. Second language learning. In: DAVIES, A; ELDER, C. (Eds.). **The handbook of applied linguistics.** United Kingdom: Blackwell, 2004. p. 501-524.

MCNAMARA, T. F. Policy and social considerations in language assessments. **Annual Review of Applied Linguistics.** n. 18, 1996, p. 304-19.

MCNAMARA, T. F.; ROEVER, C. **Language Testing: The Social Dimension**. Blackwell, Malden, Massachusetts. 2006.

MARCUZZO, P. **Ciência em debate?** Uma análise das vozes no gênero notícia da popularização científica. 2011. 173 p. Tese (Doutorado em Estudos Linguísticos). Universidade Federal de Santa Maria, Santa Maria, 2011.

MARCUZZO, P. **Uma análise do teste de suficiência em leitura em língua inglesa da UFSM**. Santa Maria: Universidade Federal de Santa Maria, 2016 (Projeto de Pesquisa).

MARCUZZO, P.; AZAMBUJA, F. O teste TOEFL-ITP dentro do Programa Ciências sem Fronteiras. **Linguagem e Ensino**, v. 20, n. 2, 333-250, 2017.

MARCUZZO, P. **Análise Crítica de Gênero de testes de inglês como língua estrangeira**. Santa Maria: Universidade Federal de Santa Maria, 2018 (Projeto de Pesquisa n. 048796).

MARCUZZO, P.; RADÜNZ, A. P. Análise crítica de gênero: uma análise de um teste de proficiência em inglês como língua estrangeira. **Fórum linguístico**, Florianópolis, v.16, n. 2, p. 3642-3654, 2019.

MARTIN, J. R. **English text: System and Structure**. Philadelphia/Amsterdam: John Benjamins, 1992.

MARTIN, J. R.; MATTHIESSEN, C. M. I. M. Modelling and mentoring: teaching and learning from home through school. In: BARRATT, L.; MAHBOOB, A. (Org.). **English in a multilingual context**. Berlin: Springer Verlag, 2012. p. 137-163.

MEDEIROS, E. B. **Provas objetivas: técnicas de construção**. Rio de Janeiro, FGV. 1977.

MESSICK, S. Validity. In: R. L. LINN (ed.), **Educational measurement**. 3 ed., New York, American Council on Education e Macmillan, p. 13-103, 1989.

MEURER, J. L. Uma dimensão crítica do estudo de gêneros textuais. Em: MEURER, J. L.; MOTTA-ROTH, D. (Org.). **Gêneros textuais e práticas discursivas: subsídios para o ensino da linguagem**. Bauru: EDUSC, p. 17-29, 2002.

MEURER, J. L. **Análise crítica de gêneros textuais: texto e contexto a partir da linguística sistêmica funcional com ênfase no componente sociológico da linguagem**. Brasília: CAPES, 2003. Projeto de Pesquisa (estágio pós-doutoral na Macquarie University, Sidney).

MEURER, J. L. Ampliando a noção de contexto na linguística sistêmico-funcional e na análise crítica do discurso. **Linguagem em Discurso**, Tubarão, v. 4, n. esp., p. 133- 157, 2004.

MEURER, J. L. Gêneros textuais na análise de Fairclough. In: MEURER, J. L.; BONINI, A.; MOTTA-ROTH, D. **Gêneros**: teorias, métodos, debates. São Paulo, Parábola Editorial, p. 81- 106, 2005.

MILLER, C. R. Genre as social action. **Quarterly Journal of Speech**, n.70, p. 151-167, 1984.

MINISTÉRIO DA EDUCAÇÃO. **Número de pós-graduandos cresce no Brasil**. Disponível em: <http://portal.mec.gov.br/setec-secretaria-de-educacao-profissional-e-tecnologica/180-estudantes-108009469/pos-graduacao-500454045/2583-sp-2021081601>. Acesso em: 26 abr. 2020.

MOTTA-ROTH, D. A dinâmica de produção de conhecimento: teorias e dados, pesquisador e pesquisados. **Revista Brasileira de Linguística Aplicada**. v. 3, n. 1, 165-177, 2003.

MOTTA-ROTH, D. Questões de metodologia em análise de gêneros. Em: KARWOSKI, A. M.; GAYDECZKA, B.; BRITO, K. S. (Org.). **Gêneros textuais**: reflexões e ensino. Palmas: Kaygangue, 2005. p. 179-202.

MOTTA-ROTH, D. Análise Crítica de Gêneros: contribuições para o ensino e a pesquisa de linguagem. **D. E. L. T. A. Documentação de Estudos em Linguística Teórica e Aplicada**, v. 24, p. 341-383, 2008.

MOTTA-ROTH, D.; HEBERLE, V. A short cartography of genre studies in Brazil. **Journal of English for Academic Purposes**, n. 19, p. 22-31, 2015.

MOTTA-ROTH, D; SCHERER, A. S. Popularização da ciência: a interdiscursividade entre ciência, pedagogia e jornalismo. **Bakhtiniana**, São Paulo, v. 11, n. 2. p. 164-189, 2016.

MUSTAFA, F; APRIADI, H. DIY: Designing a reading test as reliable as a paper-based TOEFL designed by ETS. In: **Proceedings of the 1st English Education International Conference**. Banda Aceh, Indonesia: Syiah Kuala University Press, p. 402-407, 2016.

NASCIMENTO, R. G. Análise Crítica de Gênero, planejamento de material didático e letramentos do professor de inglês como língua estrangeira/adicional. In: TOMITCH, L. M. B.; HEBERLE, V. M. (Org.). **Perspectivas atuais de aprendizagem e ensino de línguas**. 1ed. Florianópolis: LLE/PPGI/UFSC, v; 1, p. 121-152, 2017.

NEW LONDON GROUP. A pedagogy of multiliteracies: designing social futures. **Harvard Educational Review**, v. 66, n. 1, p. 60-92, 1996.

PACKER, A. Cresce a adoção do inglês entre os periódicos SciELO do Brasil [online]. **SciELO em Perspectiva**, 2016. Disponível em: <https://blog.scielo.org/blog/2016/05/10/cresce-a-adocao-do-ingles-entre-os-periodicos-scielo-do-brasil/>. Acesso em: 26 abr. 2020.

RADÜNZ, A. P; MARCUZZO, P. O TESLLE de Língua Inglesa da UFSM – um estudo sobre os tipos, enfoques e aspectos linguísticos das questões e as pontuações dos examinandos. **Revista Brasileira de Iniciação Científica**, Itapetininga, v. 4, n. 5, 2017.

ROJO, R. **Letramentos múltiplos, escola e inclusão social**. São Paulo: Parábola Editorial, 2009.

SANTOS, A. M. M. **A compreensão leitora no teste TOEFL ITP**: reflexões sobre possíveis efeitos retroativos a partir do uso das estratégias de leitura por candidatos em Porto Velho. 2016. 101 p. Dissertação (Mestrado em Letras) – Universidade Federal de Rondônia, Porto Velho. 2016.

SCARAMUCCI, M. V. R. Efeito retroativo da avaliação no ensino/aprendizagem de línguas: o estado da arte. **Trabalhos em Linguística Aplicada**, v. 46, n. 2, p. 203-226, 2004.

SCARAMUCCI, M. V. R. Avaliação da leitura em inglês como língua estrangeira e validade de construto. **Calidoscópico**, v. 7, n. 1, p. 30-48, 2009.

SILALAH, R. M. **Assessing university students' critical thinking skill by using the TOEFL ITP reading test**. *Língua Cultura*, v. 11, n. 2. p. 79-83, 2017.

SOARES, M. **Letramento** - Um tema em três gêneros. Belo Horizonte: Ceale/Autêntica, 1998.

STREET, B. **Literacy in theory and practice**. Cambridge: Cambridge University Press, 1984.

SWALES, J. M. **Genre analysis**: English in academic and research settings. Cambridge: Cambridge University Press, 1990.

SWALES, J. M. **Other floors, other voices**: a textography of a small university building. Mahwah, NJ: Lawrence Erlbaum, 1998.

SWALES, J. M. **Research genres**: exploration and applications. Cambridge: Cambridge University Press, 2004.

TAUFIQ, W.; SANTOSO, D. R.; FEDIYANTO, N. Critical Analysis on TOEFL ITP as A Language Assessment. In: **Advances in Social Science, Education and Humanities Research**, v. 125, Atlantis Press, 2018.

TESLLE. **Teste de suficiência em leitura em língua estrangeira**. Disponível em: [http://w3.ufsm.br/testedesuficiencia/images/teslle/Ingua\\_inglesa\\_amostra.pdf](http://w3.ufsm.br/testedesuficiencia/images/teslle/Ingua_inglesa_amostra.pdf). Acesso em: 29 mar. 2019.

TESLLE. **Editais**. Disponível em: <http://w3.ufsm.br/testedesuficiencia/index.php/editais>. 2019. Acesso em: 15 fev. 2020.

TESLLE. **Mini-manual do candidato**. 2018. Disponível em: [http://w3.ufsm.br/testedesuficiencia/images/Mini\\_manual\\_do\\_candidato\\_TESLLE\\_2018.pdf](http://w3.ufsm.br/testedesuficiencia/images/Mini_manual_do_candidato_TESLLE_2018.pdf). Acesso em: 06 jan. 2020.

THOMPSON, G. **Introducing functional grammar**. London: Arnold, 2004.

UNIVERSIDADE FEDERAL DE SANTA MARIA. **Resolução n. 003/10**. Disponível em: [http://w3.ufsm.br/ppggeo/images/resolucao%20003\\_2010.pdf](http://w3.ufsm.br/ppggeo/images/resolucao%20003_2010.pdf). Acesso em: 06 jan. 2020.

UNIVERSIDADE FEDERAL DE SANTA MARIA. **Resolução n. 015/2014**. Disponível em: <https://www.ufsm.br/pro-reitorias/prpgp/wp-content/uploads/sites/345/2018/04/015-2014-Regimento-da-Ps-Graduao.pdf>. Acesso em: 11 fev. 2020.

UNIVERSIDADE FEDERAL DE SANTA MARIA. **Resolução n. 018/18**. Disponível em: <https://www.ufsm.br/wp-content/uploads/2018/08/Resolu%C3%A7%C3%A3o-018-2018-UFSM.pdf>. Acesso em: 25 jan. 2020.

WATANABE, Y. Methodology in washback studies. In: CHENG, L; WATANABE, Y; CURTIS, A. (eds). **Washback in Language Testing** – Research Contexts and Methods. Lawrence Erlbaum Associates, New Jersey, 2004, p. 19-36.

WIELEWICKI, H. G. **Testagem de proficiência em leitura em inglês**: examinandos e teste como fontes de entendimento sobre esse processo. 1997. 199 p. Dissertação (Mestrado em Letras) - Universidade Federal de Santa Maria, Santa Maria, 1997.

WIGGINGS, G. P. **Introduction**: assessment and the morality of testing. In: WIGGINGS, G. P. **Assessing student performance: exploring the purpose and limits of testing**. San Francisco: Jossey-Bass Publishers, 1993. p. 1-33.

WRIGHT, S. Citizenship tests in Europe - editorial introduction. **International Journal on Multicultural Societies**, v. 10, n. 1, 1-9, 2008.

APÊNDICE A - AMOSTRA DO TESLLE DE LÍNGUA INGLESA

# TESLLE

teste de suficiência em leitura em língua estrangeira



Língua Inglesa

TESTE DE  
AMOSTRA



UFSM

DLEM  
Departamento de Letras Estrangeiras Modernas



COPERVES  
UFSM

Para responder às questões, leia o texto a seguir.

# Rigour, reliability and validity in qualitative research

T. Long and M. Johnson

This article addresses issues relating to rigour within qualitative research, beginning with the need for rigour at all in such studies. The concept of reliability is then analysed, establishing the traditional understanding of the term, and evaluating alternative terms. A similar exploration of validity and proposed alternatives follows. It is suggested that there is nothing to be gained from the use of alternative terms which, on analysis, often prove to be identical to the traditional terms of reliability and validity. Alternative or novel means of addressing these concepts in interpretive research are, however, welcomed. A review of some of the strategies available for the pursuit of reliability and validity in qualitative research is undertaken. These are clearly identified as means to establish existing criteria and are found to have variable value. © 2000 Harcourt Publishers Ltd

**Keywords:** rigour, reliability, validity, qualitative research

## THE NEED FOR RIGOUR IN RESEARCH

There is general agreement that all research studies must be open to critique and evaluation. Failure to assess the worth of a study – the soundness of its method, the accuracy of its findings, and the integrity of assumptions made or conclusions reached – could have dire consequences. Ambiguous or meaningless findings may result in wasted time and effort, while findings which are simply wrong could result in the adoption of dangerous or harmful practices. Evaluation of studies, then, is an essential pre-requisite of the application of findings. Traditionally, such evaluation has centred on assessment of reliability and validity. However, while these terms have distinct meanings which relate well to other concepts and assumptions within the logical-positivist paradigm, their use in qualitative work has been questioned. A variety of positions is to be found. These include dismissing any attempt to establish rigour; using existing terms and criteria in the traditional manner; using existing terms and criteria with modification to their interpretation; and rejecting traditional terms and criteria while substituting new terms and criteria. It is the last of these which is challenged here.

## RELIABILITY

Taking two commonly used sources for the traditional understanding of reliability, this concept can

be described as 'the consistency or constancy of a measuring instrument' (LoBiondo-Wood & Haber 1998, p. 558), or 'the degree of consistency or dependability with which an instrument measures the attribute it is designed to measure' (Polit & Hungler 1995, p. 651). Hammersley (1992, p. 67) suggests that reliability 'refers to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions.' Surprisingly, in view of their different philosophical approaches, there is little disagreement among these definitions. The first two indicate the detached nature of the researcher, while the third acknowledges the active involvement of the researcher, but all relate to confidence in data collection.

The traditional understanding of reliability focuses on standardizing data collection instruments (Mason 1996, p. 24). However, 'this is premised on the assumption that methods of data generation can be conceptualised as tools, and can be standardised, neutral and non-biased' (Mason 1996, p. 145). While this may be acceptable for quantitative methods, though Hammersley (1992) questions even this, the non-standardization of qualitative methods and the determination to seek greater validity through retention of context makes it impossible in qualitative work. Brink (1991, p. 176) proposes three tests of reliability for qualitative work, each to be used as is appropriate for specific studies. *Stability* is established when asking identical questions of an informant at different times produces consistent answers. *Consistency* refers to the

§ 1

§ 2

§ 3

**Tony Long** PhD, RN  
Senior Nursing Lecturer,  
School of Healthcare  
Studies, Baines Wing,  
University of Leeds, PO  
Box 214 Leeds LS2 9LT,  
UK

**Martin Johnson** PhD, RN  
Professor of Nursing, Dept  
of Acute & Critical Care  
Nursing, Greenbank  
Building, University of  
Central Lancashire, Preston  
PR1 2HE, UK

Correspondence to: TL at  
University of Leeds. Tel:  
+44 (0) 113 233 1297; Fax:  
+44 (0) 113 233 1204;  
E-mail: a.j.long@leeds.ac.uk

*Clinical Effectiveness in Nursing* (2000) 4, 30–37 © 2000 Harcourt Publishers Ltd

integrity of issues within a single interview or questionnaire, so that a respondent's answers on a given topic remain concordant. *Equivalence* is tested by the use of alternative forms of a question with the same meaning during a single interview, or by concurrent observation by two researchers. These approaches appear to do no more than apply standard approaches of replicability and inter-rater reliability to qualitative interviews or observation studies, seeking inappropriately to standardise highly variable data collection methods.

There seems to be a growing popular movement within qualitative circles to insist that 'dependability' is a more appropriate term than reliability for qualitative research (Sandelowski 1986; Hall & Stevens 1991; Robson 1993; Koch 1994). The origin of this movement was the work of Guba and Lincoln (1985), but the authors' thoughts were refined in their contribution of 1989. In this they explain that 'dependability is parallel to the conventional criterion of reliability, in that it is concerned with the stability of data over time' (Guba & Lincoln 1989, p. 242). However, the concern at the root of dependability is the same as that for reliability: to ensure that data collection is undertaken in a consistent manner free from undue variation which unknowingly exerts an effect on the nature of the data. A specific strategy of dependability audit is suggested by which to enhance or demonstrate dependability, but this is clearly declared by Guba and Lincoln (1989, p. 242) as a means to an end rather than a criterion in itself.

Some accounts lack any suggestion of explanation of the need for an alternative criterion. Robson (1993, p. 405), for example, provides no description or explanation of dependability other than to state that 'dependability is analogous to reliability.' The account continues to describe how dependability is established through auditing of the decision trail. As Hammersley (1992) suggests is often the case, the focus is upon means rather than criteria. A frequently quoted article by Koch (1994), which in fairness is intended to be a discussion only of the decision trail, makes no attempt to define or describe any of Guba and Lincoln's alternative criteria, but is often cited as evidence of the need for such terms (e.g. Holloway & Wheeler 1996).

The inescapable conclusion is that the terms considered above have the same essential meaning, and nothing is to be gained from clouding the issue with alternative labels for what have been argued to be identical concepts. Reliability, recognized as pertaining to the stability of data-collection measures, remains an important notion. Rather than attempting to hide behind a smokescreen of synonyms, perhaps interpretive researchers ought simply to accept that reliability is unlikely to be a demonstrable strength of their work. Although efforts may be made to enhance a study's reliability, in most cases the nature of the data and the sample make this practically hopeless.

## VALIDITY

In quantitative terms, validity is taken to mean 'the determination of whether a measurement instrument actually measures what it is purported to measure' (LoBiondo-Wood & Haber 1998, p. 561), or 'the degree to which an instrument measures what it is intended to measure' (Polit & Hungler 1995, p. 656). Hammersley (1992, p. 69) provides a qualitative perspective: 'An account is valid or true if it represents accurately those features of the phenomena that it is intended to describe, explain or theorise.' As with definitions of reliability, there seems little discrepancy between these perspectives.

While in quantitative studies validity is concerned specifically with avoidance of type I and type II errors, Silverman (1993, pp 144, 155) argues that the refutability proposed by Popper (1979) is the standard approach to testing the validity of any research findings. Hammersley (1992, p. 69) accepts that no knowledge can be counted as certain, and the best that we can do is to seek means of judging claims to knowledge in terms of their likely truth. These means are laid out as considering the plausibility of the claim, the credibility of the claim, and the weight of evidence for each of these. Silverman, noting that claims to credibility may be no more than a reflection of uncritical public ignorance, agrees that the only useful option is to consider the quality of the evidence which upholds the claims. Hammersley (1992, pp 70–72) refers to the realities of life and differing degrees of need for confidence dependent upon the significance of the claim. He suggests that some claims are within our common experience, leave little room for error on the part of the researcher, or are in no way central to the main issues of the study. These may be accepted without much concern. Other claims allow more room for misinterpretation by the researcher and therefore require stronger evidence. Where a claim is of particular significance to the study the most convincing evidence is called for. Although Silverman (1993, pp 155–156) shuns all but the most demanding of these, it is clear that Hammersley is committed to any claims (including those based on plausibility or credibility) to knowledge being supported by evidence. It is the nature of the claim and its centrality to the study which decide the degree of evidence required.

Validity is normally established through consideration of three main aspects: content validity, criterion-related validity and construct validity. The first of these depends largely on sampling and careful construction of the instrument and refers to the degree to which the entirety of the phenomenon under investigation is addressed. A sub-set of this is the weak concept of face validity which assures only that the instrument and findings appear to be thorough and accurate to reputedly knowledgeable reviewers. Criterion-related validity is concerned with comparison of the instrument and findings

§ 4

§ 5

§ 6

§ 7

§ 8

§ 9

with an established standard to determine the correlation between measured performance and actual performance. Finally, construct validity is associated with consideration of the proximity of the instrument to the construct in question.

As with reliability, those who accept the need for tests of validity in qualitative research commonly insist on the use of alternative terms, often asserting that these explain alternative concepts. Guba and Lincoln (1989, pp 236–237) adopt the term ‘credibility’ instead of validity. They argue that validity refers to the naïve reality of positivism and an attempt to establish ‘isomorphism between findings and objective reality.’ Their substitution involves replacing this with ‘isomorphism between constructed realities of respondents and the reconstructions attributed to them’ (Guba & Lincoln 1989, pp 236–237). The only difference between the terms is the presumed objective reality of positivism and the constructed realities of constructivism. The underlying concept appears to be identical: to match what is reported by the researcher to the phenomenon under investigation. As suggested by Hammersley (1992, p. 67), the notable difference lies more in the means to establish achievement of the criterion rather than in the criterion itself. However, the assertion of the need for a new criterion is accepted in a fairly unquestioning manner in a number of texts directed specifically at nurses (e.g. Holloway & Wheeler 1996).

Further claims are often made for feminist research. Hall and Stevens (1991), for example, explain ten measures of ‘adequacy’. Adequacy is preferred as a term to validity because it is held that reliability and validity cannot be separated in feminist research as they are in other approaches. In fact, they also explain and employ the term ‘dependability’, which, it is argued above, is identical to reliability, and they discuss this separately. The origin of the assumption that reliability and validity are unconnected in traditional studies is not made clear, although Hammersley (1992, p. 66) provides more argument for this. He notes that ‘the findings of a study are either valid or they are not ... The distinction between internal and external validity is fundamentally misleading.’ Adequacy is explained as follows. ‘Results are adequate if analytic interpretations fairly and accurately reflect the phenomena that investigators claim to represent’ (Hall & Stevens 1991). This appears to concur with the definitions of validity already considered, thus calling into question the assertion that ‘adequacy’ is in some way different to, or more than, validity. In seeking means to check credibility Hall and Stevens (1991) demand a depth of trust, intimacy of setting, and sensitivity to the language and life-style of the respondent in addition to sufficient length and frequency of contact in order to demonstrate ‘rapport’. The purpose of this is clearly to ensure that the full story

unfolds and thereby to ensure what is otherwise known as content validity.

‘Coherence’, say Hall and Stevens (1991), ‘can be recognised in the consistency of the whole with its constituent parts.’ Thus, the plausibility of the findings is reliant upon the interpreted, analysed data being recognisably drawn from the raw data, and is demonstrated partly by evidence of stringent efforts to ensure this. In other work (Silverman 1993), this requirement is held to be of considerable importance and is addressed, for example, by the process of analytic induction. Although not expressed by Hall and Stevens, there must also be acknowledgement of the need for the data to have been collected in a valid manner. These concerns to ensure that the findings appear believable and that they are representative of the mass of raw data (and, therefore, in turn, of the original phenomenon) are directly comparable respectively to face and content validity.

Two other criteria proposed by Hall and Stevens (1991) are more problematic: ‘complexity’ and ‘consensus’. They justifiably demand that findings are context-specific and remain sufficiently complex to reflect accurately the true nature of complex phenomena. This obviously expresses a commitment to content validity. However, they also note that ‘congruence among behavioural, verbal and affective statements of particular observations, verbal responses, and written records helps to support the presence of consensus.’ The suggestion is that credibility is enhanced when triangulated data concur, and that areas where congruence occurs should be given priority. This must automatically limit the breadth of data which is deemed to be acceptable, thereby stripping the complexity of the data through discarding contextual material. Silverman warns of this danger when utilizing triangulation. All accounts or observations could be true within their own context, and ‘counterposing different contexts ignores the context-bound and skilful character of social interaction’ (Silverman 1993, pp 156–158). Hall and Stevens (1991) admit that ‘inconsistency among participant accounts does not invalidate their perceptions, but instead illustrates the variety of women’s thoughts, actions and feelings.’ They seem to recognize the apparent disparity between complexity and consensus, but offer no solution.

‘Relationality’ is the term used by Hall and Stevens to address the group co-operation with other researchers and participants which enables larger samples, longer studies and more critical reflection. Wider sampling, larger numbers and greater depth of critical analysis of alternative explanations are principles held dear by positivists, too, and are generally sought in order to enhance content validity.

‘Reflexivity’ is the term used by Hall and Stevens (1991) for the recognition of the need to incorporate the subjective value of the researcher’s

§ 10

§ 12

§ 13

§ 11

§ 14

feelings and attitudes into consideration of the findings. A word of caution is offered by Hammersley (1992, p. 142), warning that 'there is a danger of various substantive values being smuggled in under the disguise of the formal value of reflection.' Despite this, Koch (1994) notes that giving sufficient detail of the researcher's and participants' involvement, together with existing knowledge and beliefs, allows estimation of adherence to, or departure from the theoretical construct under investigation. Such concern is usually addressed as construct validity. Although claimed as a criterion for the new concept of adequacy, reflexivity is really presented as a means to the accomplishment of the existing criterion of construct validity.

§ 15

Hall and Stevens (1991) view credibility, however, from a different perspective. For them a study 'is credible when it presents such faithful interpretations of participants' experiences that they are able to recognise them as their own.' They appear to seek credibility in the eyes only of the participants rather than external reviewers, though this is, of course, compatible with the principle proposed above by Hall and Stevens that what women say is inherently valid. For much qualitative research, there is little possibility of finding a standard against which to test the study findings and respondent validation is utilized instead. In this, the respondents themselves provide such a standard. What is apparently proposed here is precisely that, and thereby criterion-related validity is being sought.

§ 16

Two more criteria of 'honesty and mutuality' are held to relate specifically to feminist research. It is assumed that women are inherently honest and able to ignore the effects of being studied, and it is then sufficient to treat women subjects as equals for 'the subjective validity of participants' statements, affects and behaviours' to be preserved (Hall & Stevens 1991). This appears to be an unsupported assertion that women guard a self-evidently valid construct of a given phenomenon, but in any case admits concern for construct validity. This is reinforced by 'warning': since 'a study on feminist principles is adequate if the active voices of women participants are heard in the research account' (Hall & Stevens 1991). Although this is clearly no guarantee that the researcher heard, recorded or interpreted those voices accurately, it once again confirms the value of construct validity.

§ 17

Hall and Stevens' (1991) final criterion of 'relevance', which asserts that research is not valid if it fails to further the struggle for women's emancipation, is clearly not associated with the concepts of validity outlined above, but may be a specific example of what Hammersley (1992, pp 72-77) refers to as 'relevance'. While the assumptions made and terminology employed by Hall and Stevens are merely examples of alternatives to traditional terms of validity, it can be seen that there is no substantive distinction in meaning or concern.

§ 18

## THE MEANS TO ESTABLISH RIGOUR

Hammersley (1992, p. 67) suggests that there is great confusion between criteria of rigorous research and the means by which the criteria may be evaluated in qualitative research. Once disentangled the more common of these include, for reliability, audit of the decision trail and triangulation. For validity, the means include self-description and reflective journal-keeping; respondent validation; prolonged involvement; persistent observation; peer debriefing; and triangulation.

§ 19

### Self-description and reflective journal

Reflection is an essential part of qualitative research. Porter (1993) sees reflexivity as researchers reflecting on their own beliefs in the same manner as they examine those of their respondents. These beliefs and values are made explicit and taken into account so that 'rather than engaging in futile attempts to eliminate the effects of the researcher, reflexive researchers try to understand them' (Hammersley & Atkinson 1995, p. 18).

§ 20

### Respondent validation (member check)

Bank (1991) suggests the use of respondent validation (Bloor 1978) to ensure stability. Checking the results on completion of data collection or of the whole study with the respondents would, it is alleged, meet the requirements of diachronic reliability (stability over time). However, significant elements of raw data are made up of field notes, observation of non-verbal signs, and recognition of unconscious changes in tone and emphasis. These may not be acknowledged or accepted later by the respondent. Furthermore, even if validation is to be undertaken at the end of the study the time lapse involved is unlikely to be sufficient to demonstrate stability in a meaningful way. Alternatively, supposing that respondent validation were to be undertaken two or three years after a study, it might be expected to provide more security regarding the stability of the findings. There would, however, be considerable problems associated with such an attempt, such as respondent morbidity, lack of access, and alteration of the respondent's situation and views, perhaps even as a result of participation in the study. 'Member check' has been used as an alternative term for respondent validation. However, while Schein (1987, p. 51) suggests that this relates to researchers being able to pass themselves off as a member of the studied group (accepted by Bloor (1997) as one of three possibilities), the most common understanding of this would appear to be identical to the principles of respondent validation as above. The accuracy of

§ 21

the findings is checked with members of the studied group.

§ 22

Both Hammersley and Atkinson (1995, p. 227) and Mason (1996, pp 151–152) warn against placing too much faith in the results of respondent validation. The former note that 'we cannot assume that anyone is a privileged commentator on his or her own actions, in the sense that the truth of their account is guaranteed' (Hammersley & Atkinson 1995, p. 229). Participants' memory may fail, they may be unconscious of some of the non-verbal clues that they transmit which forms part of the data, or they may simply (consciously or unconsciously) deny less attractive aspects of their behaviour. Mason (1996, p. 147) adds that each individual respondent has no true insight into the experiences of other participants. Examples of all of these problems are provided by Bloor (1997, pp 41–48), who is emphatic that 'member validation is a many-splendoured thing, but it is not validation.' Hammersley and Atkinson (1995, p. 230) conclude that 'Such feedback can be highly problematic. Whether respondents are enthusiastic, indifferent, or hostile, their reactions cannot be taken as direct validation or refutation of the observer's inferences.'

§ 23

A further issue relating to this is that respondent validation is normally conducted by the researcher. It might add strength to the process if the check were to be undertaken by a third party. The difficulties associated with this would be the lack of rapport with the respondents (likely to affect their responses), and the possibility of re-interpreting the findings during the process. An intermediate position might be for the researcher to present the findings to the respondents, either by telephone follow-up to written details or by means of a simple questionnaire, and then for a third party to undertake analysis and comparison with the study findings. While this might go some way to answering accusations of researchers sitting in judgement on themselves, it would do nothing to address the main problems highlighted by Hammersley and Atkinson or by Mason. While respondent validation may be a useful addition to the means of assessing the rigour of a study, the results, whether supportive or not, must be treated with caution.

§ 24

#### Prolonged involvement and persistent observation

Kirk and Miller (1986, pp 30–31) argue that prolonged involvement in a community under research enhances sensitivity 'to discrepancies between the meanings presumed by the investigator and those understood by the target population.' Simply being in the respondent's environment enhances the likelihood of their meaning emerging and being recognised. It is a means to enhance validity, then, if the researcher can spend a significant length of time in

contact with respondents individually and with the topic generally. This allows time for emerging concepts to develop and for potential implications to be recognized. It also allows for more opportunities to test out tentative explanations. Guba and Lincoln (1989, p. 237) recommend *prolonged engagement* in order to build trust and *overcome the difficulties* presented by *perverse constructions and misinformation* on the part of respondents. *Persistent observation* enhances the effect of *the involvement* by enabling the researcher 'to identify those characteristics and elements in the situation that are most relevant to the problem or issue being pursued and to focus on them in more detail' (Guba & Lincoln 1989, p. 237).

#### Peer debriefing

Robson (1993, p. 404) briefly describes peer debriefing as 'exploring one's analysis and conclusions to a colleague or other peer on a continuous basis.' He suggests that being explicit in formulating something for presentation to a peer fosters subsequent credibility. However, this is the limit of the explanation. Similarly, Holloway and Wheeler (1996, p. 165) mention only that supervisors have a key role with research students to ensure rigour in their studies. Peer debriefing may be pursued in numerous forms. One of these is to discuss the emerging findings at intervals with knowledgeable colleagues, a second to present and defend method and findings at national research conferences, and a third to present the findings and implications to interested groups. Review with colleagues is intended to stimulate consideration and exploration of additional perspectives and explanations at various stages of the process of data collection and analysis. In particular, it is aimed at preventing premature closure of the search for meaning and patterns in the data. Presentation at research conferences is a recognized means of submitting method and findings to other researchers so as to attract and answer to critical comment. This is the process which Hammersley (1992) refers to in his discussion of levels of evidence to substantiate claims to knowledge. Presenting findings and implications to interested users offers similar opportunities but with particular emphasis on the relevance of the study.

§ 25

#### Triangulation

Triangulation may take several forms, but commonly refers to the employment of multiple data sources, data collection methods, or investigators. In general, the purpose of this would be to reduce the disadvantages inherent in the use of any single source, method or investigator. Kirk and Miller (1986, p. 30) note the usual concerns about avoiding type 1 and type 2 errors, but introduce concern for

§ 26

'type 3' errors since 'asking the wrong question actually is the source of most validity errors.' Triangulation of method is held to be an effective device to prevent this. However, Hammersley and Atkinson (1995, pp 231–232) warn against the assumption that triangulation provides evidence that some data are true while other data is false. Observations from differing sources or resulting from different methods may well be observations of a different phenomenon. This alteration may not be immediately apparent. Although the new question may not be 'the wrong question', it could well be a different one. 'If it is accepted that there are horses for courses,' suggests Bloor (1997, p. 38), 'and that, for any given topic, there will be one best method of investigation, then triangulation may be said to involve juxtaposing findings gathered by the best available method with findings generated by an inferior method.' In this, he holds that triangulation may illuminate different perspectives on the problem but does not provide any test of validity. If there is a place for triangulation, Hammersley and Atkinson (1995, p. 232) emphasize that it should be 'a matter not of checking whether data are valid, but of discovering which inferences from those data are valid.'

#### Audit of the decision trail

This technique, first proposed by Sandelowski (1986), involves the presentation of details of all sources of data, collection techniques and experiences, assumptions made, decisions taken, meanings interpreted, and influences on the researcher. This was not an original idea, however. The notion of honestly declaring the origins of value-laden concepts and publicly acknowledging potentially researcher-centred perspectives was supported by Myrdal (1970, p. 43). The purpose of declaring the decision trail is to allow others to decide on the worth of the study by following the trail taken and comparing it with their own conclusions made from the same information. It is a demonstration of the degree to which the researcher has remained true to the data and to the boundaries of the sample. There is some obvious correlation between this and replicability, the possibility of duplicating a study to determine if earlier results are repeated. The most practical difference would appear to be that the former involves theoretical activity while the latter requires active repetition of the study in question. Each tests whether assumptions are justifiable and data collection has been undertaken rigorously and reported accurately. However, Sandelowski's proposal allows for the uniqueness of each situation and recognises the nature of the claim as one of a perspective on rather than reproduction of the phenomenon. Kirk and Miller (1986, pp 55–56) include the recording of field notes in this sort of strategy. They note the importance of retaining socially

Rigour, reliability and validity in qualitative research

undesirable or irresponsible entries and distinguishing verbatim respondent items from researcher interpretations. The decision trail may be displayed in the discussion of the method used (and particularly issues relating to the sample), in the detailed exposition of data analysis, and within the discussion of findings.

#### CONCLUSION

There is a clear imperative for rigour to be pursued in qualitative research so that findings may carry conviction and strength. Reliability is a concept applicable to qualitative studies without the need for alternative terminology. While the catalogue of existing strategies for assessment of reliability may not normally be appropriate in interpretive studies, reliability in interpretive work can be assessed and presented using the existing terms but employing different means. In most cases it will be exceptionally difficult to demonstrate reliability. However, this is to be expected, as perfect validity is the sole guarantor of reliability. Validity has been found to be the quintessential element of qualitative research, sharing the same meaning and terminology as traditional approaches, but bearing especially great significance. However, alternative strategies for assessment and assurance of validity may be required for such studies. Where existing terms and concepts can be used without prejudice to data and findings from qualitative studies alternatives are superfluous and serve only to alienate those with differing priorities or stances. For clarity and utility identical terms and concepts should be addressed. The need is not for new criteria or novel terms but for different means of addressing existing criteria.

#### REFERENCES

- Bloor M 1978 On the analysis of observational data: a discussion of the worth and uses of inductive techniques and respondent validation. *Sociology* 12(3): 545–552
- Bloor M 1997 Techniques of validation in qualitative research: a critical commentary. In: Miller G, Dingwall R (eds) *Context and method in qualitative research*. Sage, London
- Brink P 1991 Issues of reliability and validity. In: Morse J (ed) *Qualitative nursing research: a contemporary dialogue*. Sage, London, pp 164–186
- Guba E, Lincoln Y 1985 *Effective Evaluation: Improving the Usefulness of Evaluation*. San Francisco, Jossey Bass
- Guba E, Lincoln Y 1989 *Fourth Generation Evaluation*. Sage, California
- Hall J, Stevens P 1991 Rigor in feminist research. *Advances in Nursing Science* 13(3): 16–29
- Hammersley M 1992 *What's Wrong with Ethnography?* Routledge, London
- Hammersley M, Atkinson P 1995 *Ethnography: Principles in Practice*, (2nd edn). Routledge, London
- Holloway I, Wheeler S 1996 *Qualitative Research for Nurses*. Blackwell Scientific, Oxford

§ 28

§ 27

[...]

Responda às questões com base no texto intitulado *Rigour, reliability and validity in qualitative research*.

01

O tema do estudo reportado no texto é

- a) valorização da pesquisa qualitativa.
- b) rigor na pesquisa qualitativa.
- c) terminologia alternativa e rigor científico.
- d) rigor nas pesquisas em geral.
- e) qualidade e rigor na pesquisa científica.

02

No parágrafo 1, qual critério NÃO é citado para definir o valor de um estudo?

- a) Integridade das afirmações.
- b) Coerência dos objetivos.
- c) Solidez do método.
- d) Precisão dos resultados.
- e) Integridade das conclusões.

03

No segmento "It is the last of these which is challenged here" (§ 1), o segmento destacado se refere a

- a) "dismissing any attempt to establish rigour" (§ 1).
- b) "using existing terms and criteria in the traditional manner" (§ 1).
- c) "using existing terms and criteria with modification to their interpretation" (§ 1).
- d) "rejecting traditional terms and criteria while substituting new terms and criteria" (§ 1).
- e) "A variety of positions is to be found" (§ 1).

04

Nos parágrafos 2 e 3, os autores do texto apresentam várias definições para confiabilidade. Associe as colunas, relacionando a definição a seu(s) respectivo(s) autor(es).

- |                                 |  |
|---------------------------------|--|
| 1. LoBiondo-Wood & Haber (1998) | ( ) Grau de consistência ou dependência com que um instrumento mede um atributo.   |
| 2. Polit & Hungler (1995)       | ( ) Grau de consistência com que ocorrências são atribuídas a mesma categoria por observadores diferentes ou pelo mesmo observador em situações distintas. |
| 3. Hammersley (1992)            | ( ) Padronização dos instrumentos de coleta de dados.  |
| 4. Mason (1996)                 | ( ) Consistência ou constância de um instrumento de medida.  |

Assinale a sequência correta.

- a) 4 - 3 - 2 - 1.
- b) 4 - 1 - 3 - 2.
- c) 2 - 3 - 4 - 1.
- d) 3 - 2 - 1 - 4.
- e) 1 - 4 - 2 - 3.

**05**

Sobre a confiabilidade, indique se os segmentos a seguir estão em acordo (A) ou desacordo (D) com o que apresenta o parágrafo 3.

- ( ) Ao se fazer perguntas idênticas a um informante em diferentes oportunidades, verifica-se a estabilidade.
- ( ) Consistência é a coerência das respostas de um mesmo informante em dois instrumentos de pesquisa diferentes.
- ( ) Ao se parafrasear uma pergunta em diferentes momentos de uma entrevista, testa-se a equivalência.

Assinale a sequência correta.

- (a) D - A - D
- (b) A - A - D
- (c) A - D - D
- (d) A - D - A
- (e) D - A - A

**06**

Sobre a validade, indique se os segmentos a seguir estão em acordo (A) ou desacordo (D) com o que apresentam os parágrafos 7 e 9.

- ( ) Na pesquisa quantitativa, a validade é o grau com que um instrumento mede aquilo que se propõe a medir.
- ( ) Na pesquisa qualitativa, a validade permite determinar até que ponto os resultados são generalizáveis a todos os fenômenos.
- ( ) A validade de conteúdo depende da amostragem e da elaboração do instrumento de coleta de dados.
- ( ) A validade de critério é obtida ao se comparar o instrumento de coleta de dados e os resultados com um padrão estabelecido a fim de aferir a correlação entre desempenho calculado e desempenho real.

Assinale a sequência correta.

- (a) A - D - D - D
- (b) A - A - D - D
- (c) D - A - A - A
- (d) D - A - A - D
- (e) A - D - A - A

**07**

Assinale a alternativa cujo fragmento sublinhado indica alto grau de certeza do autor em relação ao que afirma.

- (a) "There is general agreement that all research studies must be open to critique and evaluation" (§1).
- (b) "Failure to assess the worth of a study – the soundness of its method, the accuracy of its findings, and the integrity of assumptions made or conclusions reached – could have dire consequences" (§1).
- (c) "These may be accepted without much concern" (§8).
- (d) "The suggestion is that credibility is enhanced when triangulated data concur, and that areas where congruence occurs should be given priority" (§13).
- (e) "Alternatively, supposing that respondent validation were to be undertaken two or three years after a study, it might be expected to provide more security regarding the stability of the findings" (§21).

08

Assinale o segmento que NÃO inclui uma relação lógica de CONTRASTE entre as ideias.

- a) "The first two indicate the detached nature of the researcher, while the third acknowledges the active involvement of the researcher, but all relate to confidence in data collection" (§2).
- b) "While this may be acceptable for quantitative methods, though Hammersley (1992) questions even this, the non-standardization of qualitative methods and the determination to seek greater validity through retention of context makes it impossible in qualitative work" (§3).
- c) "Although efforts may be made to enhance a study's reliability, in most cases the nature of the data and the sample make this practically hopeless" (§6).
- d) "As suggested by Hammersley (1992, p. 67), the notable difference lies more in the means to establish achievement of the criterion rather than in the criterion itself. However, the assertion of the need for a new criterion is accepted in a fairly unquestioning manner in a number of texts directed specifically at nurses (e.g. Holloway & Wheeler 1996)" (§10).
- e) "The suggestion is that credibility is enhanced when triangulated data concur, and that areas where congruence occurs should be given priority" (§13).

09

Os segmentos destacados em "For them a study 'is credible when it presents such faithful interpretations of participants' experiences that they are able to recognize them as their own" (§16) referem-se, respectivamente, a

- a) participants – Hall and Stevens – participants – interpretations of participants' experiences.
- b) Hall and Stevens – interpretations of participants' experiences – participants – participants.
- c) Hall and Stevens – participants – participants' experiences – participants.
- d) interpretations of participants' experiences – participants – participants' experiences – participants.
- e) participants – Hall and Stevens – participants – participants' experiences.

10

O segmento destacado em "This appears to be an unsupported assertion that women guard a self-evidently valid construct of a given phenomenon" (§17) pode ser traduzido, no contexto em que é empregado, por

- a) um dado valor de um fenômeno válido e indiscutível.
- b) um construto indiscutivelmente válido de um dado fenômeno.
- c) um dado fenômeno pertencente a um construto indiscutível.
- d) um fenômeno pertencente a um construto indiscutivelmente válido.
- e) um construto válido de um dado fenômeno indiscutível.

**11**

A função do segmento destacado em "Checking the results on completion of data collection or of the whole study with the respondents would, it is alleged, meet the requirements of diachronic reliability" (§21) é

- (a) atenuar o grau de comprometimento do autor em relação ao que é dito.
- (b) estabelecer relação lógico-semântica de condição-resultado.
- (c) sinalizar alto grau de certeza do autor em relação ao que é dito.
- (d) estabelecer relação lógico-semântica de causa-consequência.
- (e) esclarecer o sentido de um segmento mencionado anteriormente no texto.

**12**

Considere as seguintes afirmações sobre a técnica denominada **Respondent validation** (§21).

I - Busca confirmar os resultados com os informantes de pesquisa.

II - Envolve problemas como dificuldade de acesso ao informante e alteração do posicionamento do informante.

III - Permite que outros pesquisadores recuperem a trajetória de pesquisa, comparem suas conclusões com as do autor da pesquisa e, assim, decidam sobre o valor do estudo.

Estão em acordo com o texto

- (a) apenas I.
- (b) apenas II e III.
- (c) apenas I e II.
- (d) apenas I e III.
- (e) I, II e III.

**13**

Qual das alternativas se refere à técnica denominada **Audit of decision trail** (§27)?

- (a) Envolve o emprego de mais de uma fonte de dados, mais de um método de coleta de dados ou mesmo mais de um investigador.
- (b) Permite que o pesquisador observe a comunidade investigada por período de tempo maior que o usual.
- (c) Prevê que o pesquisador pondere sobre suas próprias crenças à medida que verifica as crenças de seus informantes.
- (d) Envolve discutir os resultados e problemas de pesquisa com colegas que tenham conhecimento sobre o tema.
- (e) Implica o registro de detalhes sobre o processo de pesquisa e sobre as tomadas de decisão pelo pesquisador.

**14**

O texto constitui um exemplar de um(a):

- (a) Artigo de revisão.
- (b) Artigo experimental.
- (c) Editorial de periódico científico.
- (d) Resenha acadêmica.
- (e) Resumo acadêmico.

15

Relacione o nome de cada técnica (em Inglês) com o benefício que ela promove em termos de rigor na pesquisa.

- |   |   |
|---|---|
| 1. Self-description and reflective journal (§20)          | ( ) Reduz as desvantagens inerentes ao uso de uma única fonte, método ou investigador.  |
| 2. Prolonged involvement and persistent observation (§24) | ( ) Permite que o pesquisador entenda como ele próprio afeta os resultados da pesquisa ao invés de tentar eliminar sua influência.          |
| 3. Peer debriefing (§25)                                  | ( ) Reduz as discrepâncias entre os significados pressupostos pelo investigador e aqueles que são reconhecidos pela comunidade investigada. |
| 4. Triangulation (§26)                                    | ( ) Estimula o pesquisador a considerar e a explorar outras perspectivas e explicações potenciais para os dados.                            |

Assinale a sequência correta.

- (a) 3 - 4 - 1 - 2
- (b) 2 - 1 - 3 - 4
- (c) 1 - 2 - 4 - 3
- (d) 4 - 1 - 2 - 3
- (e) 4 - 3 - 2 - 1

16

Qual alternativa sintetiza a **CONCLUSÃO** (§28) a que chegam os autores do texto sobre o tema?

- (a) Para os autores, é imperativo que se busque rigor na pesquisa qualitativa, mas, para isso, é necessário propor terminologia alternativa, uma vez que a terminologia existente não parece apropriada a todas as abordagens de pesquisa.
- (b) É necessário que se busque o rigor na pesquisa qualitativa e, para isso, os autores defendem que se usem os mesmos conceitos de rigor já existentes, mas adotando-se meios diferentes para se avaliar e garantir critérios de rigor.
- (c) Para que a pesquisa qualitativa garanta o mesmo rigor adotado em pesquisas quantitativas, é necessário não apenas propor conceitos específicos sobre o que determina o rigor científico, como também estratégias para avaliá-lo e garanti-lo.
- (d) Existem termos e conceitos sobre rigor científico que se aplicam tanto na pesquisa quantitativa quanto na qualitativa e, portanto, os autores consideram supérfluo investir em novos estudos sobre rigor.
- (e) Os autores consideram que a validade é algo inerente ao rigor na pesquisa qualitativa, mas sua essência e significado não podem ser os mesmos adotados em abordagens tradicionais.

## APÊNDICE B - AMOSTRA DA SEÇÃO DE LEITURA DO TOEFL ITP

### Questions 1-7

Hotels were among the earliest facilities that bound the United States together. They were both creatures and creators of communities, as well as symptoms of the frenetic quest for community. Even in the first part of the nineteenth century, Americans were already forming the habit of gathering from all corners of the nation for both public and private, business and pleasure purposes. Conventions were the new occasions, and hotels were distinctively American facilities making conventions possible. The first national convention of a major party to choose a candidate for President (that of the National Republican party, which met on December 12, 1831, and nominated Henry Clay for President) was held in Baltimore, at a hotel that was then reputed to be the best in the country. The presence in Baltimore of Barnum's City Hotel, a six-story building with two hundred apartments, helps explain why many other early national political conventions were held there

In the longer run, too, American hotels made other national conventions not only possible but pleasant and convivial. The growing custom of regularly assembling from afar the representatives of all kinds of groups — not only for political conventions, but also for commercial, professional, learned, and avocational ones — in turn supported the multiplying hotels. By mid-twentieth century, conventions accounted for over a third of the yearly room occupancy of all hotels in the nation; about eighteen thousand different conventions were held annually with a total attendance of about ten million persons.

Nineteenth-century American hotelkeepers, who were no longer the genial, deferential “hosts” of the eighteenth-century European inn, became leading citizens. Holding a large stake in the community, they exercised power to make it prosper. As owners or managers of the local “palace of the public,” they were makers and shapers of a principal community attraction. Travelers from abroad were mildly shocked by this high social position.

1. The word “bound” in line 1 is closest in meaning to

- (A) led
- (B) protected
- (C) tied
- (D) strengthened

2. The National Republican party is mentioned in line 8 as an example of a group

- (A) from Baltimore
- (B) of learned people
- (C) owning a hotel
- (D) holding a convention

3. The word “assembling” in line 14 is closest in meaning to

- (A) announcing
- (B) motivating
- (C) gathering
- (D) contracting

4. The word “ones” in line 16 refers to

- (A) hotels
- (B) conventions
- (C) kinds
- (D) representatives

5. The word "it" in line 23 refers to

- (A) European inn
- (B) host
- (C) community
- (D) public

6. It can be inferred from the passage that early hotelkeepers in the United States were

- (A) active politicians
- (B) European immigrants
- (C) professional builders
- (D) influential citizen

7. Which of the following statements about early American hotels is NOT mentioned in the passage?

- (A) Travelers from abroad did not enjoy staying in them.
- (B) Conventions were held in them.
- (C) People used them for both business and pleasure.
- (D) They were important to the community.

**Questions 8-15**

- With Robert Laurent and William Zorach, direct carving enters into the story of modern sculpture in the United States. Direct carving — in which the sculptors themselves carve stone or wood with mallet and chisel — must be recognized as something more than just a technique. Implicit in it is an aesthetic principle as well :
- (5) that the medium has certain qualities of beauty and expressiveness with which sculptors must bring their own aesthetic sensibilities into harmony. For example, sometimes the shape or veining in a piece of stone or wood suggests, perhaps even dictates, not only the ultimate form, but even the subject matter.
- (10) The technique of direct carving was a break with the nineteenth-century tradition in which the making of a clay model was considered the creative act and the work was then turned over to studio assistants to be cast in plaster or bronze or carved in marble. Neoclassical sculptors seldom held a mallet or chisel in their own hands, readily conceding that the assistants they employed were far better than they were at carving the finished marble.
- (15) With the turn-of-the-century Crafts movement and the discovery of nontraditional sources of inspiration, such as wooden African figures and masks, there arose a new urge for hands-on, personal execution of art and an interaction with the medium. Even as early as the 1880's and 1890's, nonconformist European artists were attempting direct carving. By the second decade of the twentieth century, Americans — Laurent and Zorach most notably — had adopted it as their primary means of working.
- (20) Born in France, Robert Laurent (1890-1970) was a prodigy who received his education in the United States. In 1905 he was sent to Paris as an apprentice to an art dealer, and in the years that followed he witnessed the birth of Cubism, discovered primitive art, and learned the techniques of woodcarving from a frame maker.
- (25) Back in New York City by 1910, Laurent began carving pieces such as *The Priestess*, which reveals his fascination with African, pre-Columbian, and South Pacific art. Taking a walnut plank, the sculptor carved the expressive, stylized design. It is one of the earliest examples of direct carving in American sculpture. The plank's form dictated the rigidly frontal view and the low relief. Even its irregular shape must
- (30) have appealed to Laurent as a break with a long-standing tradition that required a sculptor to work within a perfect rectangle or square.

8. The word “medium” in line 5 could be used to refer to

- (A) stone or wood
- (B) mallet and chisel
- (C) technique
- (D) principle

9. What is one of the fundamental principles of direct carving?

- (A) A sculptor must work with talented assistants.
- (B) The subject of a sculpture should be derived from classical stories.
- (C) The material is an important element in a sculpture.
- (D) Designing a sculpture is a more creative activity than carving it.

10. The word “dictates” in line 8 is closest in meaning to

- (A) reads aloud
- (B) determines
- (C) includes
- (D) records

11. How does direct carving differ from the nineteenth-century tradition of sculpture?

- (A) Sculptors are personally involved in the carving of a piece.
- (B) Sculptors find their inspiration in neoclassical sources.
- (C) Sculptors have replaced the mallet and chisel with other tools.

(D) Sculptors receive more formal training.

12. The word “witnessed” in line 23 is closest in meaning to

- (A) influenced
- (B) studied
- (C) validated
- (D) observed

13. Where did Robert Laurent learn to carve?

- (A) New York
- (B) Africa
- (C) The South Pacific
- (D) Paris

14. The phrase “a break with” in line 30 is closest in meaning to

- (A) a destruction of
- (B) a departure from
- (C) a collapse of
- (D) a solution to

15. The piece titled *The Priestess* has all of the following characteristics EXCEPT

- (A) The design is stylized.
- (B) It is made of marble.
- (C) The carving is not deep.
- (D) It depicts the front of a person.

**Questions 16-26**

Birds that feed in flocks commonly retire together into roosts. The reasons for roosting communally are not always obvious, but there are some likely benefits. In winter especially, it is important for birds to keep warm at night and conserve precious food reserves. One way to do this is to find a sheltered roost. Solitary roosters shelter in dense vegetation or enter a cavity - horned larks dig holes in the ground and ptarmigan burrow into snow banks - but the effect of sheltering is magnified by several birds huddling together in the roosts, as wrens, swifts, brown creepers, bluebirds, and anis do. Body contact reduces the surface area exposed to the cold air, so the birds keep each other warm. Two kinglets huddling together were found to reduce their heat losses by a quarter and three together saved a third of their heat. The second possible benefit of communal roosts is that they act as information centers." During the day, parties of birds will have spread out to forage over a very large area. When they return in the evening some will have fed well, but others may have found little to eat. Some investigators have observed that when the birds set out again next morning, those birds that did not feed well on the previous day appear to follow those that did. The behavior of common and lesser kestrels may illustrate different feeding behaviors of similar birds with different roosting habits. The common kestrel hunts vertebrate animals in a small, familiar hunting ground, whereas the very similar lesser kestrel feeds on insects over a large area. The common kestrel roosts and hunts alone, but the lesser kestrel roosts and hunts in flocks, possibly so one bird can learn from others where to find insect swarms.

Finally, there is safety in numbers at communal roosts since there will always be a few birds awake at any given moment to give the alarm. But this increased protection is partially counteracted by the fact that mass roosts attract predators and are especially vulnerable if they are on the ground. Even those in trees can be attacked by birds of prey. The birds on the edge are at greatest risk since predators find it easier to catch small birds perching at the margins of the roost.

16. What does the passage mainly discuss?  
 (A) How birds find and store food  
 (B) How birds maintain body heat in the winter  
 (C) Why birds need to establish territory  
 (D) Why some species of birds nest together
17. The word "conserve" in line 3 is closest in meaning to  
 (A) retain  
 (B) watch  
 (C) locate  
 (D) share
18. Ptarmigan keep warm in the winter by  
 (A) huddling together on the ground with other birds  
 (B) building nests in trees  
 (C) burrowing into dense patches of vegetation  
 (D) digging tunnels into the snow
19. The word "magnified" in line 6 is closest in meaning to  
 (A) caused  
 (B) modified  
 (C) intensified  
 (D) combined
20. The author mentions kinglets in line 9 as an example of birds that  
 (A) protect themselves by nesting in holes

- (B) nest with other species of birds
- (C) nest together for warmth
- (D) usually feed and nest in pairs

21 The word "forage" in line 12 is closest in meaning to

- (A) fly
- (B) assemble
- (C) feed
- (D) rest

22. Which of the following statements about lesser and common kestrels is true?

- (A) The lesser kestrel and the common kestrel have similar diets.
- (B) The lesser kestrel feeds sociably but the common kestrel does not.
- (C) The common kestrel nests in larger flocks than does the lesser kestrel.
- (D) The common kestrel nests in trees; the lesser kestrel nests on the ground.

23. The word "counteracted" in line 24 is closest in meaning to

- (A) suggested
- (B) negated
- (C) measured
- (D) shielded

24. Which of the following is NOT mentioned in the passage as an advantage derived by birds that huddle together while sleeping?

- (A) Some members of the flock warn others of impending dangers.
- (B) Staying together provides a greater amount of heat for the whole flock.
- (C) Some birds in the flock function as information centers for others who are looking for food.
- (D) Several members of the flock care for the young.

25. Which of the following is a disadvantage of communal roosts that is mentioned in the passage?

- (A) Diseases easily spread among the birds.
- (B) Groups are more attractive to predators than individual birds.
- (C) Food supplies are quickly depleted.
- (D) Some birds in the group will attack the others.

26. The word "they" in line 25 refers to

- (A) a few birds
- (B) mass roosts
- (C) predators
- (D) trees

**Questions 27-38**

Perhaps the most striking quality of satiric literature is its freshness, its originality of perspective. Satire rarely offers original ideas. Instead, it presents the familiar in a new form. Satirists do not offer the world new philosophies. What they do is look at familiar conditions from a perspective that makes these conditions seem foolish, harmful, or affected. Satire jars us out of complacency into a pleasantly shocked realization that many of the values we unquestioningly accept are false. *Don Quixote* makes chivalry seem absurd; *Brave New World* ridicules the pretensions of science; *A Modest Proposal* dramatizes starvation by advocating cannibalism. None of these ideas is original. Chivalry was suspect before Cervantes, humanists objected to the claims of pure science before Aldous Huxley, and people were aware of famine before Swift. It was not the originality of the idea that made these satires popular. It was the manner of expression, the satiric method, that made them interesting and entertaining. Satires are read because they are aesthetically satisfying works of art, not because they are morally wholesome or ethically instructive. They are stimulating and refreshing because with commonsense briskness they brush away illusions and secondhand opinions. With spontaneous irreverence, satire rearranges perspectives, scrambles familiar objects into incongruous juxtaposition, and speaks in a personal idiom instead of abstract platitude.

Satire exists because there is need for it. It has lived because readers appreciate a refreshing stimulus, an irreverent reminder that they live in a world of platitudinous thinking, cheap moralizing, and foolish philosophy. Satire serves to prod people into an awareness of truth, though rarely to any action on behalf of truth. Satire tends to remind people that much of what they see, hear, and read in popular media is sanctimonious, sentimental, and only partially true. Life resembles in only a slight degree the popular image of it. Soldiers rarely hold the ideals that movies attribute to them, nor do ordinary citizens devote their lives to unselfish service of humanity. Intelligent people know these things but tend to forget them when they do not hear them expressed.

27. What does the passage mainly discuss?

- (A) Difficulties of writing satiric literature
- (B) Popular topics of satire
- (C) New philosophies emerging from satiric literature
- (D) Reasons for the popularity of satire

28. The word "realization" in line 6 is closest in meaning to

- (A) certainty
- (B) awareness
- (C) surprise
- (D) confusion

29. Why does the author mention *Don Quixote*, *Brave New World*, and *A Modest Proposal* in lines 6-8?

- (A) They are famous examples of satiric literature.
- (B) They present commonsense solutions to problems.
- (C) They are appropriate for readers of all ages.
- (D) They are books with similar stories.

30. The word "aesthetically" in line 13 is closest in meaning to

- (A) artistically
- (B) exceptionally
- (C) realistically
- (D) dependably

31. Which of the following can be found in satiric literature?

- (A) Newly emerging philosophies
  - (B) Odd combinations of objects and ideas
  - (C) Abstract discussion of morals and ethics
  - (D) Wholesome characters who are unselfish
32. According to the passage, there is a need for satire because people need to be
- (A) informed about new scientific developments
  - (B) exposed to original philosophies when they are formulated
  - (C) reminded that popular ideas are often inaccurate
  - (D) told how they can be of service to their communities
33. The word "refreshing" in line 19 is closest in meaning to
- (A) popular
  - (B) ridiculous
  - (C) meaningful
  - (D) unusual
34. The word "they" in line 22 refers to
- (A) people
  - (B) media
  - (C) ideals
  - (D) movies
35. The word "devote" in line 25 is closest in meaning to
- (A) distinguish
  - (B) feel affection
  - (C) prefer
  - (D) dedicate
36. As a result of reading satiric literature, readers will be most likely to
- (A) teach themselves to write fiction
  - (B) accept conventional points of view
  - (C) become better informed about current affairs
  - (D) reexamine their opinions and values
37. The various purposes of satire include all of the following EXCEPT
- (A) introducing readers to unfamiliar situations
  - (B) brushing away illusions
  - (C) reminding readers of the truth
  - (D) exposing false values
38. Why does the author mention "service of humanity" in line 25?
- (A) People need to be reminded to take action.
  - (B) Readers appreciate knowing about it.
  - (C) It is an ideal that is rarely achieved.
  - (D) Popular media often distort such stories.

### Questions 39-50

Galaxies are the major building blocks of the universe. A galaxy is a giant family of many millions of stars, and it is held together by its own gravitational field. Most of the material universe is organized into galaxies of stars, together with gas and dust.

- Line* There are three main types of galaxy; spiral, elliptical, and irregular. The Milky  
 (5) Way is a spiral galaxy: a flattish disc of stars with two spiral arms emerging from its central nucleus. About one-quarter of all galaxies have this shape. Spiral galaxies are well supplied with the interstellar gas in which new stars form; as the rotating spiral pattern sweeps around the galaxy it compresses gas and dust, triggering the formation of bright young stars in its arms. The elliptical galaxies have a symmetrical elliptical or  
 (10) spheroidal shape with no obvious structure. Most of their member stars are very old and since ellipticals are devoid of interstellar gas, no new stars are forming in them. The biggest and brightest galaxies in the universe are ellipticals with masses of about 10<sup>13</sup> times that of the Sun; these giants may frequently be sources of strong radio emission, in which case they are called radio galaxies. About two-thirds of all galaxies  
 (15) are elliptical. Irregular galaxies comprise about one-tenth of all galaxies and they come in many subclasses.

- Measurement in space is quite different from measurement on Earth. Some terrestrial distances can be expressed as intervals of time: the time to fly from one continent to another or the time it takes to drive to work, for example. By comparison  
 (20) with these familiar yardsticks, the distances to the galaxies are incomprehensibly large, but they too are made more manageable by using a time calibration, in this case the distance that light travels in one year. On such a scale the nearest giant spiral galaxy, the Andromeda galaxy, is two million light years away. The most distant luminous objects seen by telescopes are probably ten thousand million light years away. Their light was already halfway here before the Earth even formed. The light from the nearby Virgo galaxy set out when reptiles still dominated the animal world.

39. The word "major" in line 1 is closest in meaning to

- (A) intense
- (B) principal
- (C) huge
- (D) unique

40. What does the second paragraph mainly discuss?

- (A) The Milky Way
- (B) Major categories of galaxies
- (C) How elliptical galaxies are formed
- (D) Difference between irregular and spiral galaxies

41. The word "which" in line 7 refers to

- (A) dust
- (B) gas
- (C) pattern
- (D) galaxy

42. According to the passage, new stars are formed in spiral galaxies due to

- (A) an explosion of gas
- (B) the compression of gas and dust
- (C) the combining of old stars
- (D) strong radio emissions

43. The word "symmetrical" in line 9 is closest in meaning to

- (A) proportionally balanced
- (B) commonly seen

- (C) typical large
- (D) steadily growing

44. The word "obvious" in line 10 is closest in meaning to

- (A) discovered
- (B) apparent
- (C) understood
- (D) simplistic

45. According to the passage, which of the following is NOT true of elliptical galaxies?

- (A) They are the largest galaxies.
- (B) They mostly contain old stars.
- (C) They contain a high amount of interstellar gas.
- (D) They have a spherical shape

46. Which of the following characteristics of radio galaxies is mentioned in the passage?

- (A) They are a type of elliptical galaxy.
- (B) They are usually too small to be seen with a telescope
- (C) They are closely related to irregular galaxies.
- (D) They are not as bright as spiral galaxies.

47. What percentage of galaxies are irregular?

- (A) 10%
- (B) 25%
- (C) 50%
- (D) 75%

48. The word "they" in line 21 refers to

- (A) intervals
- (B) yardsticks
- (C) distances
- (D) galaxies

49. Why does the author mention the Virgo galaxy and the Andromeda galaxy in the third paragraph?

- (A) To describe the effect that distance has on visibility
- (B) To compare the ages of two relatively young galaxies
- (C) To emphasize the vast distances of the galaxies from Earth
- (D) To explain why certain galaxies cannot be seen by a telescope

50. The word "denominated" in line 26 is closest in meaning to

- (A) threatened
- (B) replaced
- (C) were developing in
- (D) were prevalent in