

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS SOCIAIS E HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO DE
ORGANIZAÇÕES PÚBLICAS

Ricardo da Silva Barcelos

**ANÁLISE DO INSTRUMENTO DE MEDIDA UTILIZADO NA
AUTOAVALIAÇÃO INSTITUCIONAL DA UFSM APLICADO AOS
SEUS DISCENTES, UTILIZANDO A TEORIA DA RESPOSTA AO ITEM
(TRI)**

Santa Maria, RS
2017

Ricardo da Silva Barcelos

**ANÁLISE DO INSTRUMENTO DE MEDIDA UTILIZADO NA AUTOAVALIAÇÃO
INSTITUCIONAL DA UFSM APLICADO AOS SEUS DISCENTES, UTILIZANDO A
TEORIA DA RESPOSTA AO ITEM (TRI)**

Dissertação apresentada ao Curso de Mestrado Profissional em Gestão de Organizações Públicas do Programa de Pós-Graduação em Gestão de Organizações Públicas da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Gestão de Organizações Públicas.**

Orientador: Prof. Dr. Ivan Henrique Vey

Santa Maria, RS
2017

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

BARCELOS, Ricardo da Silva

Análise do instrumento de medida utilizado na autoavaliação institucional da UFSM aplicado aos seus discentes, utilizando a Teoria da Resposta ao Item (TRI) / Ricardo da Silva Barcelos.- 2017.

135 p.; 30 cm

Orientador: Ivan Henrique Vey

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Ciências Sociais e Humanas, Programa de Pós-Graduação em Gestão de Organizações Públicas, RS, 2017

1. Avaliação de desempenho docente 2. Autoavaliação institucional 3. Teoria da Resposta ao Item I. VEY, Ivan Henrique II. Título.

© 2017

Todos os direitos autorais reservados a Ricardo da Silva Barcelos. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante citação da fonte.

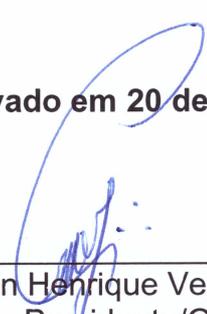
e-mail: ricardo.rsb@gmail.com

Ricardo da Silva Barcelos

**ANÁLISE DO INSTRUMENTO DE MEDIDA UTILIZADO NA AUTOAVALIAÇÃO
INSTITUCIONAL DA UFSM APLICADO AOS SEUS DISCENTES, UTILIZANDO A
TEORIA DA RESPOSTA AO ITEM (TRI)**

Dissertação apresentada ao Curso de Mestrado Profissional em Gestão de Organizações Públicas do Programa de Pós-Graduação em Gestão de Organizações Públicas da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Gestão de Organizações Públicas**.

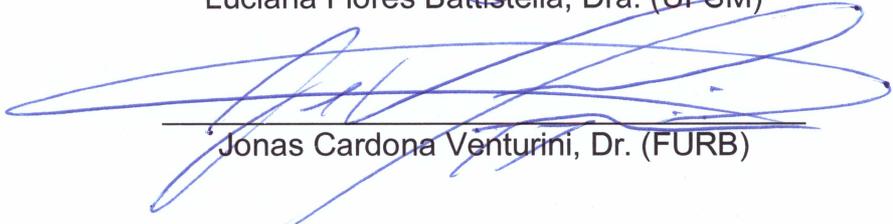
Aprovado em 20 de março de 2017.



Ivan Henrique Vey, Dr. (UFSM)
Presidente/Orientador



Luciana Flores Battistella, Dra. (UFSM)



Jonas Cardona Venturini, Dr. (FURB)

Santa Maria, RS
2017

RESUMO

ANÁLISE DO INSTRUMENTO DE MEDIDA UTILIZADO NA AUTOAVALIAÇÃO INSTITUCIONAL DA UFSM APLICADO AOS SEUS DISCENTES, UTILIZANDO A TEORIA DA RESPOSTA AO ITEM (TRI)

AUTOR: Ricardo da Silva Barcelos
ORIENTADOR: Prof. Dr. Ivan Henrique Vey

A autoavaliação institucional é um importante componente do Sistema Nacional de Avaliação da Educação Superior brasileira (SINAES), constituindo-se em um instrumento para que as Instituições de Ensino Superior identifiquem suas potencialidades e fraquezas. Uma das atividades inerentes à autoavaliação institucional é a realização da avaliação docente. Sob essa perspectiva, o presente trabalho se propôs a analisar o instrumento de medida da avaliação docente, aplicado aos discentes dos cursos de graduação da Universidade Federal de Santa Maria (UFSM). Para tanto, o estudo se baseou na técnica estatística da Teoria da Resposta ao Item (TRI) para avaliar a confiabilidade, a consistência interna e a unidimensionalidade do construto, além de estimar os parâmetros de dificuldade e discriminação do item, a partir da aplicação do Modelo Logístico de 2 Parâmetros (ML2). A TRI é um conjunto de modelos matemáticos que busca representar a probabilidade de um indivíduo responder positivamente a um item, em função dos parâmetros desse item e de sua habilidade. O tratamento e os testes estatísticos foram aplicados a uma base de dados contendo mais de 20.000 avaliações, realizadas no primeiro semestre de 2016, disponibilizada pelo Centro de Processamento de Dados da UFSM. Os resultados encontrados demonstraram que o conjunto de itens apresenta consistência interna, assim como boa qualidade de construção. O teste de dimensionalidade demonstrou que o instrumento de medida é unidimensional, medindo um único traço latente, o desempenho docente. A comparação dos parâmetros de discriminação e dificuldade estimados para a UFSM com os parâmetros estimados para os dez Centros de Ensino revelou que o comportamento dos itens é semelhante apesar das diferentes amostras. A partir dos resultados obtidos, concluiu-se que o instrumento de medida aplicado na avaliação docente pelos discentes da instituição apresentou validade estatística quando submetido aos testes da Teoria da Resposta ao Item.

Palavras-chave: Avaliação de desempenho docente. Autoavaliação institucional. Teoria da Resposta ao Item.

ABSTRACT

ANALYSIS OF THE MEASUREMENT INSTRUMENT USED IN INSTITUTIONAL SELF-EVALUATION OF UFSM APPLIED TO ITS DISCENTS, USING THE ITEM RESPONSE THEORY (IRT)

AUTHOR: Ricardo da Silva Barcelos
ADVISOR: Prof. Ivan Henrique Vey, PhD

Institutional self-assessment is an important component of the Brazilian National Higher Education Assessment System (SINAES), which consists in an instrument for higher education institutions to identify their capacities and weaknesses. One of the activities inherent to institutional self-assessment is the performance evaluation of its professors. From this perspective, this study aimed to analyze the measuring instrument used on professor performance evaluation, applied to students of undergraduate courses at the Federal University of Santa Maria (UFSM). For this, the study was based on the statistical technique of Item Response Theory (IRT) to evaluate the construct reliability, its internal consistency and unidimensionality, and to estimate the items difficulty and discrimination parameters, by using the 2-Parameter Logistic Model (ML2). The IRT is a set of mathematical models that seeks to represent the probability of an individual to answer correctly to an item, depending on the parameters of this item and its ability. The data treatment and statistical tests were applied to a database containing over 20,000 evaluations, carried out in the first half of 2016, made available by the UFSM Data Processing Center. The results showed that the set of items has internal consistency, so as good build quality. The dimensionality test showed that the measuring instrument is one-dimensional, measuring a single latent trait, the professor's performance. The comparison of the estimated discrimination and difficulty parameters for the UFSM, with the estimated parameters for the ten Study Centers revealed that the items behavior is similar despite the different samples. The results also demonstrated that the measuring instrument applied to professor performance evaluation is statistically valid when subjected to the Item Response Theory analysis.

Keywords: Professor performance evaluation. Institutional self-assessment. Item Response Theory.

LISTA DE FIGURAS

Figura 1 – Composição atual da CPA e suas Comissões Setoriais de Avaliação	47
Figura 2 – Exemplo de Curva Característica do Item (CCI)	55
Figura 3 – Curva de informação de um item	57
Figura 4 – Etapas da pesquisa	66
Figura 5 – Curvas características dos itens 1 a 8	76
Figura 6 – Curvas características dos itens 9 a 13	77
Figura 7 – Curva Característica do Item 7	79
Figura 8 – Curva de Informação do Item 7	80
Figura 9 – Curva Característica do Item 2	81
Figura 10 – Curva Característica do Item 12	82
Figura 11 – Curva de Informação Total do teste	83

LISTA DE QUADROS

Quadro 1 – Modelos matemáticos para uma única população	52
Quadro 2 – Modelos logísticos para itens dicotômicos e parâmetros componentes	53
Quadro 3 – Modelos para itens não dicotômicos e suas características	54
Quadro 4 – Métodos de estimação e seus autores	61

LISTA DE TABELAS

Tabela 1 – Número de respondentes	70
Tabela 2 – Número de avaliações por Unidade Universitária	70
Tabela 3 – Alfa de Cronbach dos itens	71
Tabela 4 – Alfa de Cronbach para o conjunto de itens	72
Tabela 5 – Variância total explicada	73
Tabela 6 – Escala de concordância	74
Tabela 7 – Correlação bisserial do conjunto de itens.	75
Tabela 8 – Parâmetros a (discriminação) e b (dificuldade) dos itens	78
Tabela 9 – Número de alunos por curso de graduação presencial – CAL	84
Tabela 10 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CAL	85
Tabela 11 – Número de alunos por curso de graduação presencial – CCNE	86
Tabela 12 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCNE	87
Tabela 13 – Número de alunos por curso de graduação presencial – CCR	88
Tabela 14 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCR	88
Tabela 15 – Número de alunos por curso de graduação presencial – CCS	90
Tabela 16 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCS	90
Tabela 17 – Número de alunos por curso de graduação presencial – CESH	91
Tabela 18 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CESH	92
Tabela 19 – Número de alunos por curso de graduação presencial – CE	93
Tabela 20 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CE	94
Tabela 21 – Número de alunos por curso de graduação presencial – CEFD	95
Tabela 22 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CEFD	95
Tabela 23 – Número de alunos por curso de graduação presencial – CT	97
Tabela 24 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CT	97
Tabela 25 – Número de alunos por curso de graduação presencial – CTISM	99
Tabela 26 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CTISM	99
Tabela 27 – Número de alunos por curso de graduação presencial – Politécnico	101
Tabela 28 – Parâmetros a (discriminação) e b (dificuldade) dos itens – Politécnico	101
Tabela 29 – Comparativo do parâmetro de discriminação por Unidade Universitária	103
Tabela 30 – Comparativo do parâmetro de dificuldade por Unidade Universitária	105

LISTA DE ABREVIATURAS E SIGLAS

ACG	Avaliação dos Cursos de Graduação
AVALIES	Avaliação das Instituições de Educação Superior
CAL	Centro de Artes e Letras
CASM	Colégio Agrícola de Santa Maria
CCI	Curva Característica do Item
CCNE	Centro de Ciências Naturais e Exatas
CCR	Centro de Ciências Rurais
CCS	Centro de Ciências da Saúde
CCSH	Centro de Ciências Sociais e Humanas
CE	Centro de Educação
CEFD	Centro de Educação Física e Desportos
CFE	Conselho Federal de Educação
CHA	Conhecimentos, Habilidades e Atitudes
CII	Curva de Informação do Item
CNRES	Comissão Nacional para Reformulação do Ensino Superior
COPLAI	Coordenadoria de Planejamento e Avaliação Institucional
CPA	Comissão Própria de Avaliação
CPD	Centro de Processamento de Dados
CSA	Comissão Setorial de Avaliação
CT	Centro de Tecnologia
CTISM	Colégio Técnico Industrial de Santa Maria
ENADE	Exame Nacional de Desempenho dos Estudantes
ENC	Exame Nacional de Cursos
ENEM	Exame Nacional do Ensino Médio
FCI	Função Característica do Item
FII	Função de Informação do Item
FIT	Função de Informação do Teste
GERES	Grupo Executivo para Reformulação do Ensino Superior
GT	Grupo de Trabalho
IES	Instituições de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Anísio Teixeira
MEC	Ministério da Educação
ML2	Modelo Logístico de Dois Parâmetros
MS-Excel	Microsoft Excel
PAIUB	Programa de Avaliação Institucional das Universidades Brasileiras
PARU	Programa de Avaliação da Reforma Universitária
PDI	Plano de Desenvolvimento Institucional
PL	Projeto de Lei
PPC	Projeto Pedagógico do Curso
PPI	Projeto Pedagógico Institucional
PPP	Projeto Político Pedagógico
PROPLAN	Pró-Reitoria de Planejamento
REUNI	Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais
SINAES	Sistema Nacional de Avaliação da Educação Superior
SISU	Sistema de Seleção Unificada
SPSS	Statistical Package for the Social Sciences

TCT	Teoria Clássica dos Testes
TRI	Teoria da Resposta ao Item
UFSM	Universidade Federal de Santa Maria

LISTA DE ANEXOS

ANEXO A – Questionário de avaliação docente pelo discente	119
ANEXO B – Curvas Características dos Itens (CCI)	122
ANEXO C – Curvas de Informação dos Itens (CII)	129

SUMÁRIO

1 INTRODUÇÃO	21
1.1 CONTEXTUALIZAÇÃO	21
1.2 TEMA E PROBLEMA DE PESQUISA	21
1.3 OBJETIVOS	24
1.4 JUSTIFICATIVA	24
1.5 ESTRUTURA DO TRABALHO	25
2 FUNDAMENTAÇÃO TEÓRICA	27
2.1 AVALIAÇÃO DA EDUCAÇÃO SUPERIOR NO BRASIL: ANTECEDENTES HISTÓRICOS	27
2.2 O SINAES	30
2.2.1 Avaliação docente no contexto do SINAES	33
2.3 AVALIAÇÃO DE DESEMPENHO	38
2.3.1 Objetivos da avaliação de desempenho	39
2.3.2 Os avaliadores de desempenho	41
2.3.3 Métodos de avaliação de desempenho	41
2.3.4 Problemas e distorções encontrados em avaliações de desempenho	43
2.3.5 Críticas à avaliação de desempenho	44
2.4 A COMISSÃO PRÓPRIA DE AVALIAÇÃO (CPA)	45
2.5 TEORIA DA RESPOSTA AO ITEM: CONCEITOS E APLICAÇÕES	48
2.5.1 Histórico	48
2.5.2 Conceitos básicos	50
2.5.3 Modelos	51
<i>2.5.3.1 O Modelo Logístico de 2 Parâmetros (ML2)</i>	<i>55</i>
2.5.4 Função de informação do item (FII)	56
2.5.5 Função de informação do teste (FIT)	58
2.5.6 Unidimensionalidade e independência local	59
2.5.7 Estimação	60
2.5.8 Vantagens e limitações da TRI	61
3 METODOLOGIA	63
3.1 CARACTERIZAÇÃO DA PESQUISA	63
3.1.1 Base filosófica	63
3.1.2 Classificações de pesquisa	64
3.1.3 Procedimentos metodológicos e técnicas de pesquisa	65
3.2 ETAPAS DA PESQUISA	66
3.3 LIMITAÇÕES DO ESTUDO	68
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS	69
4.1 TRATAMENTO DOS DADOS	69
4.2 ANÁLISE BASEADA NA TEORIA CLÁSSICA DOS TESTES (TCT)	70
4.3 ANÁLISE BASEADA NA TEORIA DA RESPOSTA AO ITEM (TRI)	74
4.3.1 Análise do resultado agregado das Unidades Universitárias	74
<i>4.3.1.1 Dicotomização dos itens</i>	<i>74</i>
<i>4.3.1.2 Análise fatorial</i>	<i>74</i>
<i>4.3.1.3 Correlação bisserial</i>	<i>75</i>
<i>4.3.1.4 Estimação dos parâmetros dos itens</i>	<i>76</i>
4.3.2 Síntese dos resultados por Unidade Universitária	84
<i>4.3.2.1 Centro de Artes e Letras – CAL</i>	<i>84</i>
<i>4.3.2.2 Centro de Ciências Naturais e Exatas – CCNE</i>	<i>86</i>
<i>4.3.2.3 Centro de Ciências Rurais – CCR</i>	<i>88</i>

4.3.2.4 Centro de Ciências da Saúde – CCS	89
4.3.2.5 Centro de Ciências Sociais e Humanas – CCSH	91
4.3.2.6 Centro de Educação – CE	93
4.3.2.7 Centro de Educação Física e Desportos – CEFD	95
4.3.2.8 Centro de Tecnologia – CT	96
4.3.2.9 Colégio Técnico Industrial – CTISM	98
4.3.2.10 Colégio Politécnico	100
4.3.3 Comparativo dos resultados obtidos por Unidade Universitária	103
5 CONCLUSÕES E RECOMENDAÇÕES	109
REFERÊNCIAS	113

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A educação, tida como motor do crescimento e desenvolvimento das potências mundiais, constitui-se em um amplo campo de estudo para temas relacionados à inovação, inclusive no que diz respeito à gestão de organizações públicas. Na esfera Federal, o ano de 2007 foi de grande valia para a educação superior, uma vez que o Decreto nº 6.096/2007 instituiu o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI. O programa teve por objetivo “criar condições para a ampliação do acesso e permanência na educação superior, no nível de graduação, pelo melhor aproveitamento da estrutura física e de recursos humanos existentes nas universidades federais” (BRASIL, 2007). A vigência do programa iniciou no ano de 2008 e se estendeu até o final de 2012.

Apesar disso, a crise político-econômica vivenciada no país nos últimos dois anos, tem resultado em contingenciamento de despesas orçamentárias. Exemplo disso foi a limitação das movimentações e empenhos relativos a dotações constantes da Lei Orçamentária de 2015, que afetou inclusive o Ministério da Educação, como a adotada pelo Decreto nº 8.581/2015, que reduziu os valores previstos para estas modalidades de R\$ 39.198.218.566,00 para R\$ 37.483.831.860,00 (BRASIL, 2015).

Nesse cenário e sob a perspectiva de que as organizações não são estanques, as instituições devem identificar novas e melhores formas de executar seus processos, e melhor atender seu público-alvo. A adoção de ferramentas de diagnóstico e de avaliação podem auxiliar o processo de tomada de decisão e promover melhorias de gestão, sendo componentes fundamentais para o desenvolvimento das organizações públicas. Dentre elas, no campo da educação superior, está a avaliação institucional, regulamentada pela Lei nº 10.861/2004, que instituiu o Sistema Nacional de Avaliação da Educação Superior (SINAES).

1.2 TEMA E PROBLEMA DE PESQUISA

Como componente do SINAES, a autoavaliação institucional está relacionada a quatro aspectos fundamentais, em consonância com a referida Lei: (1) melhoria da qualidade da educação superior; (2) orientação da expansão de sua oferta; (3)

aumento da eficácia institucional e efetividade acadêmica e social; e (4) aprofundamento dos compromissos e responsabilidades sociais das instituições (INEP, 2015).

A autoavaliação é operacionalizada internamente pelas Instituições de Ensino Superior (IES) através de grupos de trabalho criados para este fim específico, as Comissões Próprias de Avaliação (CPA). No ano de 2016, atendendo a exigências legais, a CPA da Universidade Federal de Santa Maria (UFSM), por meio das Comissões Setoriais de Avaliação, incluiu na avaliação institucional um instrumento de coleta de dados que permitiu aos alunos dos cursos de graduação avaliarem o trabalho realizado pelos docentes. Apesar de controverso, uma vez que a avaliação docente normalmente ocorre a partir da análise da produção científica, carga-horária em sala de aula e participação em projetos, e apenas para fins de progressão na carreira, a avaliação *in loco* por parte dos discentes pode se configurar em uma importante ferramenta de diagnóstico em relação às práticas pedagógicas (BARBOSA, FREIRE e CRISÓSTOMO, 2011).

Para viabilizar a avaliação docente, a Pró-Reitoria de Planejamento da UFSM, por meio da Coordenadoria de Planejamento e Avaliação Institucional (COPLAI), elaborou um instrumento de coleta de dados na forma de questionário, que foi aplicado durante o primeiro semestre letivo, especificamente no período entre 16 de junho e 16 de julho de 2016 em fase de teste. Este instrumento contou com 17 questões visando avaliar o desempenho docente, e foi respondido via Portal do Aluno, sendo disponibilizado a todos os alunos de cursos de graduação da Instituição. A aplicação em caráter oficial ocorreu entre os dias 14 de dezembro de 2016 e 6 de março de 2017, também via Portal do Aluno, não sendo obrigatória a participação dos discentes.

Conforme informações da COPLAI, foram levantados questionamentos quanto a validade do instrumento de pesquisa por parte de alguns professores, bem como sobre sua confiabilidade. A consolidação de um instrumento de medida passa invariavelmente por algum tipo de teste de validação, seja validade de conteúdo, validade de critério ou outra técnica. Esses apontamentos demonstraram a necessidade de submeter o questionário a testes de validação para, inclusive, melhorar sua aceitação por parte do segmento docente. Nesse contexto, a Teoria da Resposta ao Item (TRI) se apresenta como uma técnica estatística que pode auxiliar a validação do construto, ao testar elementos como a confiabilidade, consistência interna, correlação e parâmetros dos itens que compõem o instrumento de medida

aplicado na avaliação docente, a partir das respostas emitidas pelos avaliadores, no caso os discentes. Tendo em vista que o questionário utilizado na avaliação não foi submetido a testes de validação, a TRI se constitui em uma ferramenta útil para o preenchimento dessa lacuna.

Utilizada inicialmente em testes na área de Psicologia (PASQUALI, 2013), a TRI foi difundida quando passou a ser aplicada em testes de conhecimentos e habilidades, e em avaliações de desempenho. Trata-se de uma ferramenta usada na medição de traços latentes (como inteligência, depressão e ansiedade, por exemplo), os quais não podem ser observados diretamente. Para determinação de como se processa um traço latente em um indivíduo, a TRI gera uma escala, formada a partir dos itens elaborados com a finalidade de montar um construto.

A TRI, definida por Andrade, Tavares e Valle (2000), é um conjunto de modelos matemáticos que associam a probabilidade de um indivíduo dar certa resposta a um item com base nos parâmetros deste (discriminação e dificuldade), e em sua habilidade ou proficiência. Em primeira análise, a grande contribuição da TRI reside no fato de a técnica considerar cada item em particular e não apenas o escore total em um teste, possibilitando que as conclusões não sejam baseadas somente no teste, mas em cada item que o compõe (VEY, 2011; TAVARES, 2001).

Para se obter os resultados almejados pelo estudo, esta dissertação foi elaborada seguindo procedimentos metodológicos inspirados nos trabalhos de Vey (2011), Andrade, Tavares e Valle (2000), e outros autores com amplo conhecimento do tema. Em relação ao problema de pesquisa, este, na concepção de Gil (2010), pode ser determinado por questões de ordem prática ou intelectual. No entendimento de Marconi e Lakatos (2011), a formulação do problema não pode ser desvinculada do tema proposto, devendo esclarecer a dificuldade específica para a qual se busca solução.

Considerando a avaliação como prática necessária ao desenvolvimento como é destacado no Projeto de Avaliação Institucional da Universidade Federal de Santa Maria, elaborado em junho de 2008, que preceitua que o tema deve ser encarado como exigência de controle de qualidade e deve possuir caráter permanente, este trabalho buscou responder ao seguinte problema de pesquisa: qual a confiabilidade e validade do instrumento de medida aplicado aos discentes na avaliação dos docentes da UFSM?

1.3 OBJETIVOS

Os objetivos da pesquisa foram traçados a partir do detalhamento do problema de pesquisa, e levando em consideração seus fatores determinantes. Este trabalho teve por objetivo geral analisar o comportamento do conjunto de itens utilizado no instrumento de medida da avaliação docente pelos discentes de graduação da UFSM, utilizando a Teoria da Resposta ao Item.

Os objetivos específicos surgiram a partir de desdobramentos do objetivo geral, com vistas a responder ao problema de pesquisa. No caso do presente trabalho, os objetivos específicos delimitados foram:

- Verificar a consistência interna, confiabilidade e dimensionalidade do instrumento de medida aplicado aos discentes;
- Realizar a análise fatorial de informação plena para verificar a unidimensionalidade do instrumento de medida;
- Analisar a correlação bisserial dos itens que compuseram o instrumento de medida;
- Verificar o comportamento dos parâmetros de discriminação (parâmetro a) e dificuldade (parâmetro b) dos itens para a Instituição;
- Analisar o comportamento dos parâmetros a e b nos Centros de Ensino, comparando-o com o resultado geral;
- Propor melhorias para o instrumento de avaliação, caso sejam verificadas inconsistências.

1.4 JUSTIFICATIVA

A autoavaliação institucional, um dos temas centrais do presente estudo e parte integrante do SINAES, é fundamental para as instituições de ensino superior identificarem suas potencialidades e fraquezas, e dessa forma tomarem melhores decisões acerca de seus rumos. A falta de uma cultura gerencial no serviço público, apesar de alguma evolução ser percebida a partir dos anos 90 com introdução dos conceitos de *New Public Management*, é fator que dificulta o crescimento das instituições públicas de maneira geral. Nesse sentido, ações de melhoria serão sempre bem vistas. Além disso, não raro são observadas na mídia críticas à

ineficiência do setor público em suas diversas esferas de atuação, levantando-se assim a bandeira das privatizações.

A motivação para o presente trabalho envolveu questões de ordem teórica e prática. De ordem teórica, a relevância do estudo envolve a avaliação do desempenho docente na percepção dos discentes, uma vez que estes podem ser considerados o público-alvo da atividade de docência. Portanto, é fundamental que existam mecanismos de avaliação que permitam a aferição do desempenho docente, e que estes mecanismos sejam testados, validados e adequados para que o desempenho seja devidamente medido, um dos objetivos deste trabalho.

Como relevância prática, espera-se que o trabalho contribua para a difusão do tema e para a realização de novas pesquisas e, principalmente, que os resultados aqui obtidos, assim como os resultados da avaliação sejam efetivamente utilizados pelos gestores da Instituição na busca constante pela excelência do ensino. Espera-se ainda que o estudo contribua para o aperfeiçoamento do processo de avaliação, ao buscar a validação do instrumento de medida através da Teoria da Resposta ao Item. Apesar de a avaliação institucional possuir uma sistemática bem estruturada na Universidade, o fato de não ter sido notada a utilização de seus resultados nos processos decisórios se caracterizou como uma das principais motivações para o desenvolvimento do trabalho. Sendo assim, do ponto de vista gerencial, este estudo poderá contribuir para que gestores repensem suas ações, a partir da análise criteriosa que a técnica estatística da Teoria da Resposta ao Item possibilita.

1.5 ESTRUTURA DO TRABALHO

Este trabalho apresenta-se estruturado em cinco capítulos, organizados de acordo com uma estrutura lógica visando abordar todos os aspectos da pesquisa. O primeiro capítulo contém a Introdução, em que há uma contextualização do tema, a delimitação do problema de pesquisa, os objetivos (geral e específicos), e a justificativa e relevância da realização da pesquisa.

O segundo capítulo contém o referencial teórico utilizado para embasar a pesquisa, apresentando os conceitos e teorias centrais que deram suporte ao estudo. O capítulo foi dividido em cinco seções visando sua melhor compreensão, e abordou os postulados teóricos dos seguintes temas: a avaliação da educação superior no Brasil; o Sistema Nacional de Avaliação da Educação Superior (SINAES); a avaliação

de desempenho e seus princípios; a Comissão Própria de Avaliação (CPA); e a Teoria da Resposta ao Item (TRI).

O terceiro capítulo apresenta os aspectos metodológicos, além da caracterização da pesquisa e dos procedimentos adotados na tentativa de atingir os objetivos pretendidos e responder ao problema de pesquisa.

No quarto capítulo é apresentada a análise dos dados e discussão dos resultados, em que os resultados obtidos são interpretados inicialmente à luz da Teoria Clássica dos Testes (TCT) e, em seguida, a partir da fundamentação proposta pela TRI.

O quinto capítulo expõe as considerações finais, além das limitações encontradas para realização do estudo e das sugestões para pesquisas futuras em relação ao tema.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo abordou os temas principais, relevantes aos objetivos propostos pelo presente trabalho: a avaliação do ensino superior brasileiro através do Sistema Nacional de Avaliação da Educação Superior (SINAES), sua evolução, diretrizes e orientações, a avaliação de desempenho, e os conceitos e aplicações da Teoria da Resposta ao Item.

2.1 AVALIAÇÃO DA EDUCAÇÃO SUPERIOR NO BRASIL: ANTECEDENTES HISTÓRICOS

A avaliação da educação superior brasileira deu seus primeiros passos oficiais, ao que tudo indica, na década de 60, durante o regime militar. Expõe Zandavalli (2009), que uma das primeiras iniciativas nesse sentido ficou conhecida como Plano Atcon, elaborado pelo teórico norte-americano Rudolph Atcon no ano de 1965, com o objetivo de reestruturar o modelo de educação superior brasileiro, baseado no modelo vigente à época nos Estados Unidos. O documento final desta assessoria prestada por Atcon, a convite do Ministério da Educação e Cultura, foi intitulado de “Rumo à Reformulação Estrutural da Universidade Brasileira”, baseando-se principalmente nos pressupostos de racionalidade, eficiência e eficácia das instituições (ZANDAVALLI, 2009). Independente do mérito de o documento defender uma forte participação privada nas universidades, a autora destaca sua contribuição para a evolução dos processos avaliativos:

O Plano Atcon carrega grande importância para a área de avaliação, pois constitui um dos primeiros processos avaliativos oficiais da estrutura da universidade brasileira. E a aplicação de suas sugestões também pressupõe processos avaliativos, na medida em que busca controlar a utilização de recursos e gerar uma correspondência entre investimentos e resultados (ZANDAVALLI, 2009, p. 392).

Em 1968, conforme relata Fávero (2006), a urgência de uma reestruturação para fortalecimento das universidades a partir dos debates originados pelas mobilizações estudantis, passa a exigir do Governo atitudes para combater alguns dos problemas enfrentados pelas instituições. Uma das medidas tomadas foi a edição do Decreto 62.937, de 02 de julho de 1968, que criou um Grupo de Trabalho (GT)

encarregado de estudar e propor medidas a serem tomadas para superar o que se chamou de “crise da universidade”. Uma das importantes conclusões do relatório do GT foi de que o sistema de faculdades tradicionais como as universidades estavam organizadas não era eficaz ou capaz de suportar a expansão do ensino de acordo com as demandas nacionais, configurando-se em um entrave para o desenvolvimento científico e para a inovação. Destaca a autora:

Entre as medidas propostas pela Reforma, com o intuito de aumentar a eficiência e a produtividade da universidade, sobressaem: o sistema departamental, o vestibular unificado, o ciclo básico, o sistema de créditos e a matrícula por disciplina, bem como a carreira do magistério e a pós-graduação (FÁVERO, 2006, p. 34).

Apesar de passados quase 48 anos desta reforma, muitas Instituições de Ensino Superior (IES), como é o caso da UFSM em seu campus sede, ainda adota este sistema, à exceção do ingresso por meio do vestibular unificado, pois desde o primeiro semestre de 2015 a Instituição adotou a nota Exame Nacional do Ensino Médio (ENEM) como critério classificatório para ingresso por meio do Sistema de Seleção Unificada (SISU). O SISU é um portal eletrônico disponibilizado pelo Ministério da Educação (MEC), através do qual as instituições públicas de ensino oferecem suas vagas aos candidatos participantes do ENEM. Zandavalli (2009) reforça em seu trabalho que o relatório gerado pelo GT foi fundamental para subsidiar a criação do Projeto de Lei (PL) nº 32, que culminou com a redação da Lei nº 5.540/68, conhecida como Lei da Reforma Universitária. A referida lei teve boa parte de seus dispositivos revogados a partir da edição da Lei nº 9.394, de 20 de dezembro de 1996 – Lei de Diretrizes e Bases da Educação Brasileira.

A década seguinte, destaca Zandavalli (2009), não trouxe contribuições significativas em termos de inovação avaliativa, pois foi caracterizada pela “operacionalização gradativa das proposições realizadas nos anos de 1960” (ZANDAVALLI, 2009, p. 401).

Na década de 80, ao final do governo militar, surge o Programa de Avaliação da Reforma Universitária (PARU), por iniciativa do Conselho Federal de Educação (CFE). Destacam Barreyro e Rothen (2008) que o PARU se desdobrou a partir de discussões internas do CFE, decorrentes das greves ocorridas nas universidades federais. No entendimento dos autores,

O PARU pretendia realizar investigação sistemática da realidade, por meio de estudos, pesquisas e debates, tanto na implementação das propostas da reforma universitária de 1986 quanto das particularidades institucionais e regionais (...). Para concreção dos seus objetivos, considerava necessária a participação, especialmente, da comunidade acadêmica, realizado 'um processo de reflexão sobre sua própria prática' e de outros setores externos às instituições, sobretudo, na expressão de demandas e expectativas quanto ao papel da universidade e sua organização (BARREYRO e ROTHEN, 2008, p. 133-134).

O comentário anterior sobre a participação da comunidade acadêmica na definição dos novos rumos que as universidades deveriam tomar denota uma preocupação incipiente com a autoavaliação e o autoconhecimento das instituições para a formalização de uma proposta de reforma. Na mesma linha de pensamento, Zandavalli (2009) destaca que o programa se baseou em levantamentos e análises das condições das instituições, “desenvolvidos a partir de pesquisas, e como estratégias e etapas de desenvolvimento trabalhou com estudos de base, que deveriam permitir a avaliação comparativa de todo o sistema de educação superior, bem como examinar particularidades desse nível de ensino” (ZANDAVALLI, 2009, p. 404). Barreyro e Rothen (2008) propõem uma reflexão sobre o papel da avaliação dentro do PARU, ao afirmarem que esta foi trabalhada como uma forma de se conhecer a realidade das instituições, além de obter dados sobre sua situação e facilitar a reflexão sobre a prática, recorrendo à avaliação institucional.

Na década de 80 também foram significativos os trabalhos e relatórios apresentados pela Comissão Nacional para Reformulação do Ensino Superior (CNRES), e pelo Grupo Executivo para Reformulação do Ensino Superior (GERES), em 1985 e 1986, respectivamente (BARREYRO e ROTHEN, 2008; ZANDAVALLI, 2009). Barbosa, Freire e Crisóstomo (2011) destacam os pontos comuns a ambos os relatórios, como, por exemplo, uma proposta de implantação de sistemas de avaliação e valorização do desempenho das instituições, e a latente necessidade de maior eficiência no processo de financiamento, através da melhoria da gestão dos recursos públicos. Esta até hoje é uma questão central e difícil de ser pacificada se a gestão da instituição não tiver objetivos e direcionamento claro, pois um aumento nas fontes de financiamento pouco ou de nada adiantará se o gasto público não estiver alinhado com a visão de futuro da universidade.

No que concerne à avaliação, Barbosa, Freire e Crisóstomo (2011) destacam os aspectos que o relatório da CNRES indica como passíveis de monitoramento: “avaliação de cursos, avaliação de alunos, avaliação de professores, avaliação

didático-pedagógica do ensino, avaliação de servidores técnico-administrativos, e, avaliação das carreiras” (BARBOSA, FREIRE e CRISÓSTOMO, 2011, p. 321). Além disso, os autores assinalam que o relatório discorre sobre modalidades de avaliação, a saber: a autoavaliação, a avaliação governamental, a avaliação por parte da comunidade e as avaliações independentes. Pode-se inferir que, neste momento, surge o embrião do que hoje é o Sistema Nacional de Avaliação da Educação Superior – o SINAES.

Sobre o GERES, os autores expõem que o relatório reforçou o aspecto da necessidade de avaliação dos cursos de graduação, de forma semelhante ao que já era feito com a pós-graduação, com a proposta de se avaliar o desempenho institucional e a qualidade dos cursos (BARBOSA, FREIRE e CRISÓSTOMO, 2011). Complementando este raciocínio, cabe a transcrição do pensamento de Zandavalli (2009) sobre a importância da avaliação, a partir da análise do relatório emitido pelo GERES:

A importância da avaliação é reforçada no item “Autonomia e Avaliação”. O Grupo demonstra a vinculação entre os dois aspectos, na medida em que considera papel do Estado a regulação da qualidade da educação oferecida pelas IES públicas, pois é financiada com recursos públicos. Observa que é insuficiente o controle realizado pelo Estado, já que limitado ao controle burocrático dos meios da máquina pública e não de seus resultados efetivos. Como decorrência da proposta de ampliação da autonomia da gestão universitária, o GERES “[...] propõe que o controle social seja feito a partir de um sistema de avaliação de desempenho” (ZANDAVALLI, 2009, p. 415).

Apesar da evolução na sistemática de avaliação no transcorrer destes mais de 50 anos, mesmo o SINAES, sistema vigente abordado na próxima seção, ainda carece de estudos e regulamentações que o tornem mais aplicado e efetivo.

2.2 O SINAES

O Sistema Nacional de Avaliação da Educação Superior foi inicialmente previsto no inciso VIII do art. 9º da Lei nº 9.394 de 20 de dezembro de 1996, e criado pela Lei nº 10.861, de 14 de abril de 2004. Com o objetivo de “assegurar processo nacional de avaliação das instituições de educação superior, dos cursos de graduação e do desempenho acadêmico de seus estudantes” (BRASIL, 2004), através da avaliação dos aspectos referentes a estes três pilares, o SINAES evoluiu a partir do

Programa de Avaliação Institucional das Universidades Brasileiras (PAIUB) e do Exame Nacional de Cursos – ENC (SILVA e GOMES, 2011).

Nesse sentido, Ribeiro (2015) discorre sobre o funcionamento da avaliação nos três níveis, expondo brevemente a atuação de cada um deles. A avaliação dos cursos é realizada por especialistas designados pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP) especificamente para este fim, em três momentos distintos: autorização, reconhecimento e renovação de reconhecimento.

A primeira delas ocorre quando uma instituição solicita autorização ao Ministério da Educação para a criação de um curso, sendo realizada *in loco* por dois avaliadores sorteados entre os cadastrados no Banco Nacional de Avaliadores. São consideradas na avaliação a organização didático-pedagógica, a composição do pessoal (docente e técnico-administrativo) e as instalações físicas (INEP, 2015). Havendo a autorização para abertura, a avaliação para reconhecimento deverá ser solicitada pela instituição a partir do momento em que a primeira turma inicia a segunda metade do curso. Nesta fase ocorre também a avaliação do corpo discente, além dos itens já citados na etapa de abertura.

Por fim, a avaliação para renovação do reconhecimento deverá ocorrer de acordo com o ciclo trienal do SINAES, em que a cada ano um ou mais grupos de áreas de formação são avaliados.

São considerados para fins de avaliação, além do tripé educacional de ensino, pesquisa e extensão, também a responsabilidade social, o desempenho acadêmico, a gestão da instituição, seu corpo docente, instalações e outros critérios.

As finalidades do SINAES são definidas já no parágrafo primeiro do artigo inicial da Lei nº 10.861/2004, assim dispostas:

§1º O SINAES tem por finalidades a melhoria da qualidade da educação superior, a orientação da expansão da sua oferta, o aumento permanente da sua eficácia institucional e efetividade acadêmica e social e, especialmente, a promoção do aprofundamento dos compromissos e responsabilidades sociais das instituições de educação superior, por meio da valorização de sua missão pública, da promoção dos valores democráticos, do respeito à diferença e à diversidade, da afirmação da autonomia e da identidade institucional (BRASIL, 2004).

Pode-se notar com base no texto do parágrafo que o SINAES se propõe não só a traçar um perfil institucional a partir de suas avaliações, mas que também

apresenta caráter gerencial ao elencar como uma de suas finalidades a de orientar a expansão da oferta de educação superior.

Cabe mencionar que, no inciso I do art. 2º, a Lei nº 10.861/2004 traz um importante dispositivo para os fins do presente estudo. Nele fica estabelecido que o SINAES, ao promover a avaliação das instituições, dos cursos e do desempenho estudantil, deverá assegurar:

I – **avaliação institucional, interna e externa**, contemplando a análise global e integrada das dimensões, estruturas, relações, compromisso social, atividades, finalidades e responsabilidades sociais das instituições de educação superior e de seus cursos (BRASIL, 2004, grifo nosso).

Merece comentário o fato de que as avaliações, independentemente do nível em que ocorram, devem ser utilizadas como instrumentos de apoio a decisão. Nesta linha, Aves e Souza (2013) argumentam que “essas avaliações devem gerar ações internas nas Instituições de Ensino Superior com o intuito de qualificar seus serviços, instalações e profissionais” (AVES e SOUZA, 2013, p. 232). Também nesse sentido, Peixoto (2009) apresenta similar perspectiva da análise dos componentes da avaliação, ao sugerir que as contribuições não se restringem somente aos resultados da avaliação, mas também sobre o conhecimento que se tem sobre práticas avaliativas para que as próprias práticas possam ser qualificadas. Em outros termos, a avaliação não é concluída após a tabulação de dados e confecção de relatórios, de fato ela deve seguir sendo alvo de melhorias em seu próprio processo, pois a tendência é que a fidedignidade da avaliação aumente com sua maturidade. Concluindo, as avaliações devem ser vistas como norteadoras do processo de implementação de políticas de melhoria das condições educacionais, tanto no âmbito das IES, quanto na formulação de macropolíticas por parte do Ministério da Educação (MEC).

Aprofundando-se o estudo da Lei nº 10.861/2004, outro dispositivo que adiante irá orientar metodologicamente este trabalho diz respeito às dimensões de avaliação integrantes do SINAES, que está contido no art. 3º da referida Lei, e que merece transcrição em sua íntegra:

Art. 3º – A avaliação das instituições de educação superior terá por objetivo identificar o seu perfil e o significado de sua atuação, por meio de suas atividades, cursos, programas, projetos e setores, considerando as diferentes dimensões institucionais, dentre elas obrigatoriamente as seguintes:

- I – a missão e o plano de desenvolvimento institucional;
- II – a política para o ensino, a pesquisa, a pós-graduação, a extensão e as respectivas formas de operacionalização, incluídos os procedimentos para estímulo à produção acadêmica, as bolsas de pesquisa, de monitoria e demais modalidades;
- III – a responsabilidade social da instituição, considerada especialmente no que se refere à sua contribuição em relação à inclusão social, ao desenvolvimento econômico e social, à defesa do meio ambiente, da memória cultural, da produção artística e do patrimônio cultural;
- IV – a comunicação com a sociedade;
- V – as políticas de pessoal, as carreiras do corpo docente e do corpo técnico-administrativo, seu aperfeiçoamento, desenvolvimento profissional e suas condições de trabalho;
- VI – organização e gestão da instituição, especialmente o funcionamento e representatividade dos colegiados, sua independência e autonomia na relação com a mantenedora, e a participação dos segmentos da comunidade universitária nos processos decisórios;
- VII – infraestrutura física, especialmente a de ensino e de pesquisa, biblioteca, recursos de informação e comunicação;
- VIII – planejamento e avaliação, especialmente os processos, resultados e eficácia da autoavaliação institucional;
- IX – políticas de atendimento aos estudantes;
- X – sustentabilidade financeira, tendo em vista o significado social da continuidade dos compromissos na oferta da educação superior (BRASIL, 2004).

Levando-se em consideração as dez dimensões elencadas no art. 3º, caberá à Comissão Própria de Avaliação Institucional (CPA) de cada instituição a operacionalização deste processo. Dentre algumas atividades comuns estão a divulgação, elaboração de instrumento de pesquisa, levantamento e análise dos dados que irão compor o Relatório Anual de Autoavaliação. Em relação ao tema, Peixoto (2009) discorre sobre a importância da autoavaliação enquanto processo “cíclico, criativo e renovador de análise e síntese das dimensões da IES, permitindo a reanálise das prioridades estabelecidas no projeto político institucional” (PEIXOTO, 2009, p. 13), promovendo a aproximação da comunidade acadêmica na proposição de novas alternativas e elaboração de novas práticas. Essa afirmação corrobora o disposto no § 1º do artigo inicial da Lei do SINAES.

2.2.1 Avaliação docente no contexto do SINAES

Os dezesseis artigos da Lei do SINAES norteiam os princípios e fatores que devem ser considerados na promoção da avaliação da educação superior, em sua maioria em linhas gerais. A avaliação de desempenho docente não é tratada diretamente na referida lei, porém, subentende-se sua necessidade, tendo em vista

que o docente é o principal componente na difusão do conhecimento e de posturas críticas e reflexivas.

Alguns indícios disso são encontrados, por exemplo, no inciso V do art. 3º da Lei nº 10.861/2004, que trata das políticas de pessoal e das carreiras do corpo docente e técnico-administrativo, com vistas a propiciar seu aperfeiçoamento, desenvolvimento profissional e condições de trabalho (BRASIL, 2004). Reforçando este posicionamento, merece transcrição na íntegra o art. 4º da referida lei:

Art. 4º – A avaliação dos cursos de graduação tem por objetivo identificar as condições de ensino oferecidas aos estudantes, em especial as relativas ao perfil do corpo docente, às instalações físicas e à organização didático-pedagógica (BRASIL, 2004).

A partir desse trecho, pode-se inferir que ao tratar das condições de ensino, em especial relativas ao perfil do corpo docente, tem-se um primeiro vislumbre da importância da avaliação deste grupo. Nesse sentido, Rodrigues (2012) em trabalho intitulado “Políticas de Avaliação Docente: tendências e estratégias”, elenca algumas das finalidades da avaliação docente. Dentre elas estão o objetivo de melhorar o desempenho dos professores, assim como as práticas e procedimentos, e também como forma de prestação pública de contas, ou seja, transparência. Apesar de o trabalho da autora ter seu foco voltado para a educação básica, os pressupostos citados anteriormente também se aplicam à educação superior.

No contexto da avaliação docente, Castanheira e Ceroni (2007) trazem uma reflexão que reforça a importância do assunto, ao afirmarem:

Neste contexto, evidencia-se a necessidade de os educadores considerarem a avaliação importante meio de diagnóstico de seu trabalho, bem como a elaboração de instrumentos mais diretos de aferição da qualidade de ensino e aprendizagem do aluno que quer se formar (CASTANHEIRA e CERONI, 2007, p. 722).

A afirmação das autoras evidencia a necessidade de adoção de formas alternativas de avaliação docente. Este, normalmente tem sua atuação avaliada indiretamente pelo desempenho dos acadêmicos em exames como o Exame Nacional de Desempenho dos Estudantes (ENADE), além da avaliação de sua produção científica e outros requisitos específicos de cada instituição para fins de progressão na carreira. O desenvolvimento de ferramentas de avaliação mais efetivas, que permitam ao docente um *feedback* de sua atuação nas atividades de ensino, pesquisa

e extensão, vai ao encontro das dimensões do SINAES e de suas finalidades, principalmente a de propiciar a melhoria da qualidade da educação superior, como corrobora o posicionamento de Embiruçu, Fontes e Almeida (2010):

A avaliação do desempenho docente em Instituições de Ensino Superior (IES), de forma similar à avaliação do desempenho de qualquer profissional nos diversos campos de atuação e nos diversos sistemas e processos produtivos, é de fundamental importância para o desenvolvimento profissional dos docentes e para o crescimento institucional, ajudando a garantir que as metas qualitativas e quantitativas esperadas pela sociedade possam ser alcançadas (EMBIRUÇU, FONTES e ALMEIDA, 2010, p. 797).

Reforçando a importância da avaliação para melhoria da qualidade do ensino, em especial a avaliação docente por parte do discente, Boclin (2004) destaca que “no processo de ensino tudo acontece na razão direta das relações entre professores e alunos e cujo palco e cenário vem a ser as salas de aulas, laboratórios, bibliotecas...” (BOCLIN, 2004, p. 959). Ainda conforme o autor, o melhor entendimento destas relações e os efeitos sobre a qualidade do ensino e da aprendizagem, é ao mesmo tempo fundamental e desafiador para a continuidade das ações educativas, que se encontram em constante transformação. Seguindo o mesmo raciocínio, Sousa e Andriola (2014) apregoam que a avaliação docente deve interagir fortemente com a gestão de recursos humanos, principalmente em razão da lógica de formação continuada, uma vez que somente a concessão de oportunidades para que o professor possa se qualificar e adquirir novas competências poderá contribuir para a melhoria da aprendizagem dos alunos.

Dentre os benefícios que podem advir do processo, além de identificar o perfil do docente, bem como os aspectos didáticos ou acadêmicos nos quais ele deva se aperfeiçoar, Embiruçu, Fontes e Almeida (2010) apontam: fundamentação da avaliação em estágios probatórios, avaliação de desempenho de departamentos e Unidades Universitárias; e avaliação de desempenho da instituição como um todo. Nesse sentido, entende-se que a avaliação de desempenho docente impacta toda a instituição, dada sua importância.

No sentido de reafirmar a relevância da avaliação docente não só enquanto instrumento para aferição de desempenho, mas também com vistas ao desenvolvimento e melhoria dos processos de aprendizagem, Hoffmann (2003) enuncia:

O ato de avaliar se faz presente em todos os momentos da vida humana [...] ele também está presente em todos os momentos vividos em sala de aula. [...] A avaliação efetiva vai ocorrer durante o processo, nas relações dinâmicas de sala de aula, que orientam as tomadas de decisões frequentes, relacionadas ao tratamento do conteúdo e à melhor forma de compreensão e produção do conhecimento pelo aluno. [...] É nas relações cotidianas entre professor e aluno que vai se dar a aprendizagem (HOFFMANN, 2003, apud CASTANHEIRA e CERONI, 2007, p. 723).

Depreende-se da afirmação anterior a importância da construção de instrumentos de avaliação docente por parte de seu público-alvo direto. Castanheira e Ceroni (2007) destacam ainda que o levantamento da opinião discente em relação às atividades didático-pedagógicas dos professores é um dos componentes do processo de avaliação institucional. Destacam, porém, se tratar de tarefa complexa, “pois ao professor cabe, por vezes, não satisfazer seus alunos e suas expectativas imediatistas” (CASTANHEIRA e CERONI, 2007, p. 724). Tem-se aqui um indício de que a construção de um instrumento adequado deve atentar para possíveis distorções que podem ocorrer, principalmente decorrente de respondentes que interpretem a avaliação docente como forma de retaliação a determinado professor.

Com o objetivo de minimizar as distorções supracitadas, uma outra questão a ser considerada quando da implementação de um processo de avaliação docente por parte dos discentes é explorada por Castanheira e Ceroni (2007), ao tratar da etapa de sensibilização à criação do novo procedimento. Para as autoras, existe o risco de a avaliação não ser adequada caso os agentes envolvidos não tenham consciências de seus papéis, e da importância da avaliação:

Para que se possa implantar um processo de avaliação docente com sucesso é preciso realizar uma sensibilização de todos os envolvidos. Deve estar claro para o corpo docente que o objetivo da avaliação é a melhoria e não a punição, e, para os alunos a importância de sua opinião (que deve ser a mais justa possível e não um ajuste de contas), para que possam contribuir para a melhoria de sua formação (CASTANHEIRA e CERONI, 2007, p. 725).

O mesmo entendimento é apresentado por Boclin (2004), ao destacar três aspectos a serem considerados quando da promoção de um processo de avaliação docente por parte dos discentes: 1) a participação dos alunos no processo; 2) a adoção de sua metodologia pelo projeto político-pedagógico dos cursos; e 3) sua utilização como instrumento para elevação da qualidade dos cursos de graduação.

Uma segunda leitura sobre estes posicionamentos permite-nos inferir que o processo de avaliação deve ser entendido como um instrumento em que docentes e

alunos colaboram entre si, portanto, são parceiros, com uma finalidade comum, que é a melhoria da qualidade do ensino. No que diz respeito aos papéis de cada um na relação docente-discente, conforme apregoam Castanheira e Ceroni (2007), é fundamental o entendimento por parte de ambos que o processo de avaliação em nenhum momento tem caráter punitivo, mas tem seu foco voltado para o crescimento da instituição em termos de qualidade de ensino. Analogamente, Ribeiro (2015) aponta dois aspectos que justificariam tal processo de avaliação, não se restringindo apenas ao objetivo da melhoria da qualidade do ensino através do aperfeiçoamento de práticas didático-pedagógicas, mas incluindo uma variável econômico-financeira, por assim dizer:

Talvez seja justo pensar que o consenso em torno da avaliação foi produto, de um lado, da capacidade de demonstração da utilidade dos seus resultados, seja para conhecimento da realidade do sistema de ensino superior, seja para prestar contas à sociedade, seja para subsidiar decisões políticas e administrativas neste campo. E, de outro lado, pressionado pela reforma do Estado, a escassez de recursos para financiamento dos altos custos gerados pelos sistemas de ensino superior exige, cada vez mais, como contrapartida, a qualidade dos cursos e a eficiência das universidades (RIBEIRO, 2015, p. 59).

Cabe ressaltar que a qualidade da avaliação docente dependerá em grande parte das características do instrumento a ser utilizado para tal fim. Avrichir e Dewes (2006) destacam que os questionários que compõem o processo avaliativo, a serem preenchidos pelos alunos, devem possuir algumas características que facilitem a interpretação das assertivas e minimizem a arbitrariedade das respostas dadas, levando em consideração as categorias identificadas na literatura. São exemplos das categorias identificadas: despertar o interesse, conhecimento do assunto, clareza, preparação e organização do curso, etc.

No entendimento de Murray (1983, apud AVRICHIR e DEWES, 2006), a terminologia a ser utilizada quando da construção das assertivas deve priorizar expressões que façam referência a comportamento, ao invés de termos genéricos e abstratos. Como exemplo, Murray sugere que uma assertiva que se refira à clareza contenha expressões como “o professor usou exemplos concretos” ou então “definiu os termos empregados”.

Pode-se observar, a partir do exposto na literatura, que a qualidade da avaliação de desempenho depende em boa medida do instrumento utilizado para tal

fim. A próxima seção proporciona uma revisão dos aspectos principais referentes ao tema da avaliação de desempenho.

2.3 AVALIAÇÃO DE DESEMPENHO

A avaliação de desempenho é um subsistema dentro da área de recursos humanos/gestão de pessoas, assim como os subsistemas de recrutamento e seleção, remuneração e benefícios, treinamento e desenvolvimento, entre outros. Antes de abordar a avaliação de desempenho propriamente dita, é importante o entendimento da conceituação de desempenho humano. Nesse sentido, uma abrangente definição inicial é proposta por Marras (2011):

Desempenho humano é o ato ou efeito de cumprir ou executar determinada missão ou meta previamente traçada. É diretamente proporcional a duas condições do ser humano: o “querer fazer”, que explicita o desejo endógeno de realizar (a motivação), e o “saber fazer”, isto é, a condição cognitiva e experiencial que possibilita ao indivíduo realizar com eficácia alguma coisa (MARRAS, 2011, p. 165).

Aspecto digno de nota referente ao exposto é a perspectiva inerente ao indivíduo do “querer fazer” e do “saber fazer”, que se constituem em dois grandes desafios ligados à gestão de pessoas nas organizações. O primeiro reside na dificuldade de se manter os indivíduos motivados, na constante tentativa de alinhar objetivos pessoais e organizacionais. Já o segundo, diz respeito à necessidade de se adotar uma estratégia clara de treinamento e capacitação dos funcionários. Nesse aspecto, a avaliação de desempenho pode colaborar sobremaneira.

Fidelis e Banov (2007) apresentam entendimento semelhante e complementar ao conceito, ao afirmarem:

Entende-se por desempenho humano o ato de cumprir ou executar determinada meta previamente traçada, e por avaliação de desempenho o processo que mede o desempenho do funcionário nas metas para ele traçadas. O desempenho do funcionário é o grau em que ele alcança os requisitos de seu trabalho (FIDELIS e BANOV, 2007, p. 101).

Aprofundando um pouco mais o conceito, os autores expõem que “a avaliação do desempenho dentro das organizações é o processo que visa medir o quanto um empregado é capaz de produzir dentro de uma meta estabelecida” (FIDELIS e BANOV, 2007, p. 101). Dessler (2003) amplia esta definição ao tratar o tema como a

avaliação da relação entre o desempenho atual ou passado, e os padrões de desempenho estabelecidos.

Concepção semelhante é apresentada por Chiavenato (2009), ao afirmar que a avaliação é a apreciação sistemática do desempenho de uma pessoa no cargo que ela ocupa, e o seu potencial para desenvolvimento futuro. Aqui cabe a consideração de que o processo de avaliação de desempenho não se limita ao preenchimento de um simples instrumento de coleta de informações. Tão importante quanto, “a avaliação de desempenho também presume que os padrões de desempenho foram estabelecidos e que você dará *feedback* ao funcionário para ajudá-lo a eliminar deficiências de desempenho ou a manter seu bom nível” (DESSLER, 2003, p. 172). Nesse sentido, o autor faz referência à importância da preparação de uma entrevista de avaliação, em que supervisor e funcionário irão revisar a avaliação e planejar formas de suprir as deficiências e reforçar os aspectos positivos.

A condução da avaliação deve ser estrategicamente preparada tendo em vista o fato de poucas pessoas gostarem de receber ou dar *feedback* negativo (FEDOR e PARSONS, 1996, apud DESSLER, 2003). Entendimento similar é partilhado por França (2009), ao afirmar:

Toda avaliação de desempenho possui seu retorno, que consiste na divulgação das informações para o funcionário. Os resultados da avaliação de desempenho certamente têm papel determinante na motivação dos funcionários. O objetivo da avaliação de desempenho é também proporcionar o crescimento e o desenvolvimento da pessoa que fica sabendo como está se saindo no trabalho (FRANÇA, 2009, p. 117).

Verifica-se, a partir do exposto, que a efetividade de um processo de avaliação de desempenho depende da comunicação dos resultados ao seu maior interessado, no caso, o avaliado. Parte-se do pressuposto que, se não for dada ciência ao avaliado, dificilmente sua postura e seu desempenho sofrerão mudanças significativas com o passar do tempo.

2.3.1 Objetivos da avaliação de desempenho

Os objetivos da avaliação de desempenho foram se alterando com o passar do tempo, conforme aponta Marras (2011). Na concepção do autor, a avaliação de desempenho foi inicialmente criada para acompanhar o desenvolvimento cognitivos

dos empregados, e mensurar o seu nível de conhecimentos, habilidades e atitudes (CHA), tendo seus resultados utilizados para fins de definições de políticas de treinamento e desenvolvimento. Todavia, destaca o autor que com o passar do tempo, os programas de avaliação de desempenho tiveram suas dimensões e abrangência ampliadas “para fazer frente às necessidades surgidas no âmbito do controle administrativo” (MARRAS, 2011, p. 166).

Em termos mais específicos, a avaliação de desempenho nas organizações tem por objetivo verificar o resultado de um investimento realizado numa trajetória profissional pelo retorno recebido pela organização. Além disso, é um instrumento que serve para: 1) verificar a contribuição (do funcionário ou grupo) nos resultados da empresa; 2) identificar quem precisa de treinamento; 3) descobrir novos talentos; 4) administrar cargos e salários; 5) avaliar o sistema de remuneração; 6) promover pessoas; 7) elaborar programas de mérito; 8) promover transferências; 9) fornecer *feedback* aos empregados; 10) efetuar demissões (FIDELIS e BANOVA, 2007; MARRAS, 2011).

A concepção de Dessler (2003) sobre os objetivos da avaliação de desempenho são mais abrangentes e estão conjugadas em três bases. Em primeiro lugar, avaliar o desempenho facilita a tomada de decisão no que diz respeito a promoções; em segundo lugar, fornece informações para que chefes e funcionários desenvolvam um plano para corrigir eventuais deficiências observadas. Por fim, as avaliações podem ser utilizadas para planejamento ou reestruturação de carreiras, a partir da análise das forças e fraquezas encontradas.

Mesmo entendimento é proposto por Lacombe (2011) ao discorrer que as avaliações se destinam, em sua finalidade precípua, a melhorar o desempenho dos avaliados em suas atividades. No entanto, não devem ser desconsiderados seus objetivos secundários, como o fato de dar indicações sobre necessidades de treinamento, identificar problemas na estrutura organizacional, e mesmo para proporcionar *feedback* para processos de seleção.

Na tentativa de facilitar o estabelecimento das metas, Fidelis e Banov (2007) propõem quatro formas distintas pelas quais a avaliação pode ser balizada: 1) pelos resultados; 2) pelo conhecimento; 3) pelas habilidades; e 4) pelo comportamento.

Para o estudo em questão, os objetivos gerais da avaliação de desempenho elencados pelos autores, devem ser analisados com muito cuidado por estarmos tratando de uma instituição pública, na qual existe a figura da estabilidade. Além disso,

devem ser consideradas as peculiaridades inerentes à atividade docente no momento de definir quais serão os objetivos específicos da avaliação, bem como das dimensões que serão avaliadas. Uma vez que as práticas docentes são de natureza essencialmente intelectual (à exceção de algumas atividades rotineiras como correção de provas, divulgação de notas e controle de frequência acadêmica), a definição das metas que serão avaliadas se constitui em uma atividade bastante complexa.

2.3.2 Os avaliadores de desempenho

São as organizações, de acordo com a sua história e cultura organizacional, que decidem quem deve avaliar o desempenho de seus empregados (FIDELIS e BANOVA, 2007, p. 103). No caso de instituições públicas, em especial a UFSM, objeto deste trabalho, sabe-se que, apesar de gozarem de autonomia administrativa, são regidas por um arcabouço legal que limita sua atuação perante à sociedade. Exemplo disso é a lei do SINAES, já citada.

No caso deste estudo, tendo em vista que as análises foram efetuadas em relação aos resultados obtidos pela aplicação aos discentes de um instrumento de medida para avaliar o desempenho docente, tem-se então como avaliadores de desempenho os próprios discentes, e a Comissão Própria de Avaliação como órgão avaliador.

2.3.3 Métodos de avaliação de desempenho

As avaliações de desempenho podem ser informais, em que consistem em perguntar ao superior imediato ou aos pares como é visto o desempenho ou o trabalho de determinada pessoa na organização, ou então formal, quando são utilizados instrumentos construídos para esta finalidade (FIDELIS e BANOVA, 2007).

Dentre os métodos mais utilizados nas avaliações de desempenho formais, estão:

- 1 – Escalas gráficas: estabelece variáveis ou indicadores de desempenho, como, por exemplo, conhecimento, assiduidade, proatividade, etc. Estas são divididas em graus (excelente, ótimo, regular, etc.) aos quais são atribuídos pontos. Ao final, o resultado obtido pelo funcionário é comparado à meta estabelecida ou à média do grupo (FIDELIS e BANOVA, 2007);

2 – Incidentes críticos: leva em consideração os pontos fortes e fracos de quem está sendo avaliado, e duas séries de afirmativas (uma positiva e outra negativa). Os aspectos considerados fortes e fracos são atribuídos pelo responsável pelo setor (FIDELIS e BANOVA, 2007). Já na concepção de Dessler (2003), este método consiste no levantamento e registro de exemplos de comportamentos de trabalho excepcionalmente bons ou indesejáveis de um funcionário, e periodicamente estes registros são revistos e revisados juntamente ao avaliado. O autor destaca ainda que este método normalmente é utilizado para suplementar um método de classificação, pois “garante que o supervisor pense sobre a avaliação do funcionário durante o ano todo, já que os incidentes precisam ser registrados (DESSLER, 2003, p. 178).

3 – Avaliação 360 graus: forma de avaliação bastante utilizada, em que o trabalhador é avaliado por todos aqueles com quem tem contato: subordinados, superiores, clientes internos e externos, fornecedores, etc. (FIDELIS e BANOVA, 2007). Este método, conforme pontua Dessler (2003), em geral é utilizado para desenvolvimentos do funcionário, e não para fins de aumento salarial.

4 – Avaliação participativa por objetivos (APO): conta com a participação do funcionário e de seu superior imediato, e os objetivos são formulados por ambos. Ao fechamento de determinado período, os objetivos são avaliados conjuntamente e o processo é reiniciado, reciclando-se os objetivos da avaliação participativa.

5 – Método de classificação alternada: baseia-se na definição dos critérios a serem avaliados, em que cada funcionário é ranqueado do que apresenta melhor desempenho ao que apresenta o pior, com base no respectivo critério (DESSLER, 2003). O autor destaca que a utilidade do método reside no fato de que normalmente é mais fácil distinguir entre os melhores e os piores em um critério do que atribuir uma pontuação a cada funcionário.

6 – Comparação entre pares: também conhecido como comparação binária (GIL, 2011), este método se baseia em uma matriz em que o desempenho de cada funcionário é comparado ao de seus pares, atribuindo-se um sinal de positivo quando o avaliado apresenta desempenho superior em um quesito, em relação a cada um dos colegas, e um sinal negativo quando tem desempenho avaliado como inferior. Ao final, da mesma forma como no método de classificação alternada, tem-se um ranking do melhor ao pior classificado, no critério avaliado (DESSLER, 2003).

Além dos métodos supracitados, Gil (2011) aponta alguns outros menos utilizados, mas também encontrados na literatura, como a pesquisa de campo, o

método de frases descritivas – similar ao método da escolha forçada –, e a autoavaliação. Com relação a este último, o autor explicita que “só apresenta validade quando aplicado a grupos com notório grau de maturidade profissional” (GIL, 2011, p. 161).

2.3.4 Problemas e distorções encontrados em avaliações de desempenho

As avaliações de desempenho podem apresentar uma série de problemas ou distorções que decorrem em parte da subjetividade inerente ao processo na relação avaliador – avaliado, e também em parte por problemas na construção do instrumento de avaliação, principalmente no tocante a escalas gráficas (DESSLER, 2003). Dentre os problemas mais comuns encontrados nas avaliações de desempenho, o autor elenca:

1 – Padrões obscuros: ocorre quando uma escala de avaliação é aberta demais à interpretação, desfavorecendo a uniformidade de percepções em relação a determinado critério de avaliação. Nesse aspecto, Dessler (2003) destaca que “a melhor maneira de retificar esse problema é desenvolver e incluir frases descritivas que definam cada aspecto e grau de mérito” (DESSLER, 2003, p. 183).

2 – Efeito halo: trata-se da tendência de um aspecto ou critério avaliado influenciar a avaliação dos demais, positiva ou negativamente. Por exemplo, uma característica pessoal como dificuldade em se relacionar pode afetar negativamente a avaliação de produtividade, se o avaliador não tiver plena consciência da existência desse problema.

3 – Erro de tendência central: bastante comum, consiste em avaliar todos os funcionários pela média (DESSLER, 2003), evitando-se atribuir as notas mais baixas ou mais altas da escala. Uma das formas de se contornar este problema é a utilização de métodos de classificação alternada.

4 – Brandura ou rigor: semelhante ao erro de tendência central, mas com a concentração das respostas nos extremos superior ou inferior. De maneira análoga, uma forma de se evitar este tipo de situação é a utilização de métodos de classificação.

5 – Parcialidade: no entendimento de Dessler (2003), trata-se da “tendência de permitir que diferenças individuais como idade, raça e sexo afetem a avaliação que os funcionários recebem” (DESSLER, 2003, p. 183).

2.3.5 Críticas à avaliação de desempenho

O subsistema de avaliação de desempenho, como qualquer área de estudo, é alvo de críticas quanto à sua aplicabilidade e utilização. Nesse contexto, no entanto, Gil (2011) destaca que estas críticas se justificam em função de equívocos incorridos na implantação do sistema, e de sua inadequação metodológica em relação às novas realidades empresariais. O autor elenca as seis críticas principais relacionadas à avaliação de desempenho:

1 – Definição inadequada de objetivos: diz respeito à falta de clareza dos propósitos da avaliação. Como destaca o autor, para muitos avaliadores, “o preenchimento dos formulários visa apenas satisfazer a área de recursos humanos, pois não conseguem perceber consequências concretas tais como programas de treinamento, remanejamentos, promoções, etc.” (GIL, 2011, p. 150).

2 – Baixo nível de envolvimento da alta administração: em geral, o envolvimento da alta administração é maior nas decisões concernentes a áreas como Marketing e Produção (GIL, 2011). Nesse aspecto, o papel fundamental da área de recursos humanos é demonstrar a importância das pessoas na organização para a gestão do negócio.

3 – Despreparo para administrar pessoas: nesse quesito, uma das características fundamentais ao avaliador ou gestor deve ser sua capacidade de comunicação principalmente para fins de treinamento, resolução de conflitos e para fornecer *feedback*. Gil (2011) assinala que é imprescindível que seja ministrado treinamento específico quando da implementação de sistemas de avaliação de desempenho.

4 – Avaliação circunscrita a um dia: esta crítica refere-se à falta de continuidade do procedimento de avaliação de desempenho, e com a tendência de os avaliadores serem influenciados pelos acontecimentos mais recentes (GIL, 2011). Uma forma de minimizar este efeito é transformar a avaliação em um processo contínuo, ou com periodicidade bastante definida.

5 – Baixo nível de participação das chefias: segundo Gil (2011), as chefias tendem a apresentar baixo comprometimento com a avaliação de desempenho, tendo sua participação restrita ao preenchimento de formulários.

6 – Exclusão dos empregados do processo de planejamento: esse aspecto revela a importância da participação conjunta de chefias, empregados e da área de recursos humanos, para construção conjunta do processo de avaliação.

Marras (2011) aborda a vulnerabilidade da avaliação de desempenho principalmente no tocante à subjetividade inerente ao processo. Afirma o autor:

Na prática, é preciso destacar e reconhecer que esse instrumento carrega no seu processo uma dose forte de subjetividade, que o investe de uma ambiguidade realmente discutível. Essa ambiguidade, contudo, deve ser reconhecida apenas em certos tipos de metodologia utilizados na aplicação do instrumento (MARRAS, 2011, p. 167).

Levando em consideração algumas das críticas citadas, Lacombe (2011) ressalta a importância da existência de planos formais de avaliação como forma de minimizar possíveis aspectos negativos. Nesse sentido o autor afirma que, mesmo que nenhuma avaliação seja totalmente objetiva, “a existência de um plano formal de avaliação deve conduzir a avaliações menos subjetivas e menos distorcidas do que as realizadas sem nenhuma estruturação” (LACOMBE, 2011, p. 46).

2.4 A COMISSÃO PRÓPRIA DE AVALIAÇÃO (CPA)

A Comissão Própria de Avaliação (CPA) tem por finalidade a realização da avaliação periódica das atividades desenvolvidas na Universidade, tendo a assessoria e o apoio da Coordenadoria de Planejamento e Avaliação Institucional (COPLAI). Esta, por sua vez, além de promover e dar suporte à CPA em suas atividades, desenvolve atividades ligadas aos processos de avaliação institucional, e é parte da estrutura organizacional da Pró-Reitoria de Planejamento (PROPLAN). Pode-se afirmar que, no âmbito das Instituições de Ensino Superior (IES), a CPA é um órgão avaliador de desempenho.

A base legal para a criação e funcionamento das Comissões Próprias de Avaliação está concentrada fundamentalmente no *caput* do art. 11 da Lei nº 10.861/2004 e em seus incisos I e II, conforme segue:

Art. 11 – Cada instituição de ensino superior, pública ou privada, constituirá Comissão Própria de Avaliação - CPA, no prazo de 60 (sessenta) dias, a contar da publicação desta Lei, com as atribuições de condução dos processos de avaliação internos da instituição, de sistematização e de

prestação das informações solicitadas pelo INEP, obedecidas as seguintes diretrizes:

I – constituição por ato do dirigente máximo da instituição de ensino superior, ou por previsão no seu próprio estatuto ou regimento, assegurada a participação de todos os segmentos da comunidade universitária e da sociedade civil organizada, e vedada a composição que privilegie a maioria absoluta de um dos segmentos;

II – atuação autônoma em relação a conselhos e demais órgãos colegiados existentes na instituição de educação superior (BRASIL, 2004).

Uma observação feita por Peixoto (2009) e que vai ao encontro do disposto no artigo em questão, no sentido de que a legislação atribui bastante liberdade às instituições para a constituição de suas CPA, indica que sua formação pode resultar em uma diversidade de composições e características, peculiares a cada instituição de ensino. Nesse sentido, Almeida Junior (2005) discorre sobre o surgimento e constituição das CPA, com base no disciplinado nos artigos de 9 a 17 da Portaria nº 2.051/2004, que regulamenta o SINAES. A referida Portaria dispõe que os procedimentos avaliativos do SINAES são integrados por três instrumentos distintos, em consonância com o inciso I do art. 2º da Lei nº 10.861/2004, a saber: Avaliação das Instituições de Educação Superior (AVALIES); Avaliação dos Cursos de Graduação (ACG); e Exame Nacional de Avaliação de Desempenhos dos Estudantes (ENADE) (ALMEIDA JUNIOR, 2005).

A Avalies é subdividida em duas etapas, como já mencionado neste estudo, sendo uma externa – realizada por comissão designada para este fim pelo INEP – e uma interna, a cargo das CPA. Na concepção do autor, apesar da autonomia garantida às instituições para condução da autoavaliação, ao menos três etapas devem ser seguidas (ALMEIDA JUNIOR, 2005):

1 – Preparação: momento em que há a constituição da CPA, sensibilização dos agentes e elaboração do projeto de avaliação;

2 – Desenvolvimento: início da atuação, incluindo a coleta de dados e informações, e as análises cabíveis; e,

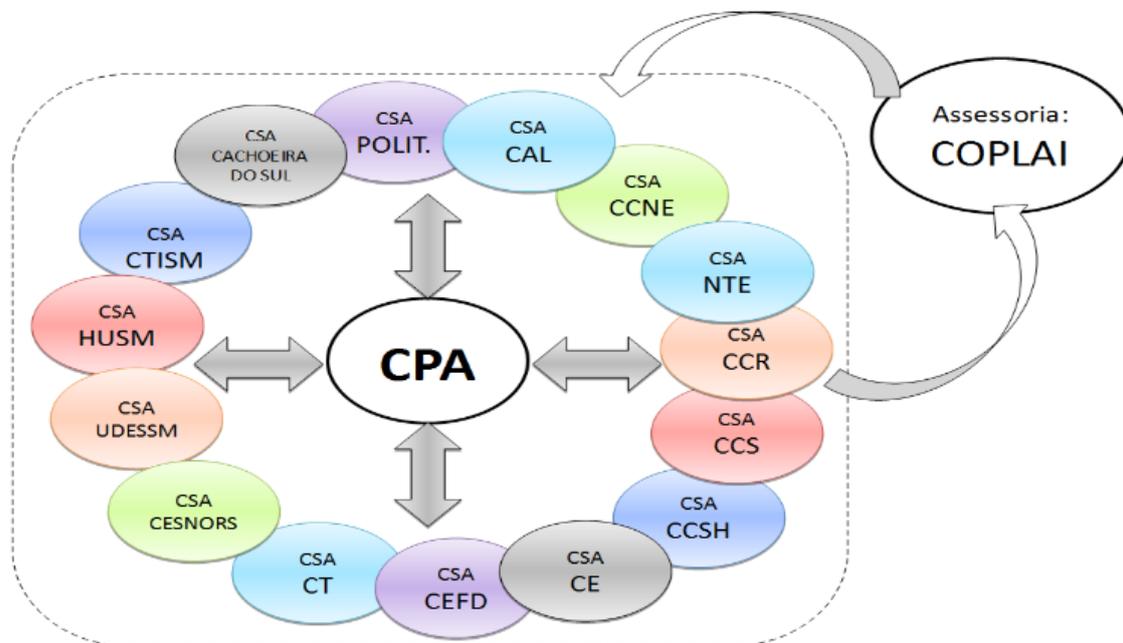
3 – Consolidação: confecção de relatório, ações de divulgação e balanço crítico.

No âmbito da Universidade Federal de Santa Maria, a CPA e suas atribuições estão contidas no capítulo 6 do Plano de Desenvolvimento Institucional (PDI), tendo seu Regimento Interno sido instituído pela Resolução nº 009/2015 da UFSM. De acordo com o PDI, o processo de autoavaliação é desenvolvido “em colaboração com as Comissões Setoriais de Avaliação das Unidades de Ensino, contando com a

assessoria da Coordenadoria de Planejamento e Avaliação Institucional (COPLAI)” (UFSM, 2011). A coordenadoria mencionada está vinculada à Pró-Reitoria de Planejamento da Universidade. O capítulo expõe a metodologia que irá conduzir o processo de avaliação, abordando cada uma de suas etapas desde o planejamento das ações, passando pela fase de sensibilização, até a elaboração do relatório final e sua divulgação. Ao final deste ciclo ou diagnóstico, os resultados obtidos serão utilizados para formulação e implementação de ações e melhorias, a serem registradas no próximo PDI, no Projeto Pedagógico Institucional (PPI), nos Projetos Pedagógicos dos Cursos (PPC) e em planos de gestão, como forma de apoiar a tomada de decisão (UFSM, 2011).

Em função principalmente da estrutura organizacional e das dimensões da UFSM, a concepção de sua CPA incluiu a criação das Comissões Setoriais de Avaliação (CSA), totalizando quinze unidades em 2015. A Figura 1 ilustra a atual composição da CPA.

Figura 1 – Composição atual da CPA e suas Comissões Setoriais de Avaliação



Fonte: <http://coral.ufsm.br/cpa/index.php/apresentacao/estrutura>.

O Projeto de Avaliação Institucional da UFSM foi concluído no mês de junho de 2008, e se constitui no documento basilar da condução de todo o processo na Universidade. Nele estão sinalizados os objetivos geral e específicos da avaliação, as

etapas metodológicas e outras orientações relativas à elaboração do relatório final, da divulgação dos resultados e do balanço crítico. Historicamente, desde a década de 80 a Instituição já se utilizava de alguns mecanismos de autoavaliação, em busca de alternativas para adoção de um Sistema de Avaliação de Desempenho.

No atual Projeto de Avaliação estão relatadas algumas das ferramentas utilizadas ao longo dos anos (UFSM, 2008): Diagnóstico de desempenho da Universidade Federal de Santa Maria (1980); Proposta de Avaliação da Universidade Federal de Santa Maria (1986); Indicadores de Desempenho dos Departamentos da UFSM (1989). No entanto, a institucionalização da Avaliação parece ter ocorrido somente a partir do ano de 1993, com o Projeto Político Pedagógico aprovado no mesmo ano. O texto base do documento dá ênfase à avaliação institucional “como um trabalho permanente de reflexão do fazer Universitário, sendo a primeira condição para identificar os desafios necessários à formulação de diretrizes para o Ensino, a Pesquisa e a Extensão” (UFSM, 2008).

2.5 TEORIA DA RESPOSTA AO ITEM: CONCEITOS E APLICAÇÕES

A avaliação, seja ela no âmbito de organizações públicas ou privadas, é uma atividade fundamental para se iniciar qualquer processo de melhoria. Se não houver diagnóstico de uma situação ou problema, por si só dificilmente haverá evolução e crescimento. Nesse sentido, o uso da Teoria da Resposta ao Item (TRI) pode contribuir sobremaneira para formulação de diagnósticos e auxiliar a avaliação de maneira complementar a Teoria Clássica dos Testes (TCT), tanto no sentido de validar instrumentos de pesquisa quanto na interpretação dos resultados obtidos.

2.5.1 Histórico

A Teoria da Resposta ao Item, conforme apontado por Pasquali (1996), deu seus primeiros passos com os trabalhos de Lord nos anos de 1952 e 1953 nos Estados Unidos, e de Rasch (1960), sendo utilizada inicialmente em testes de desempenho e aptidão. Conforme aponta Vey (2011), estes autores propuseram os primeiros modelos estatísticos paramétricos para itens de testes, ao associar a probabilidade de determinada resposta (certa ou errada) a uma variável latente (um traço não diretamente observável). A variável ou traço latente é interpretada “com a

proficiência ou habilidade dos respondentes em um contexto de testes para avaliação educacional ou psicométrica (VEY, 2011, p. 99). Pasquali (1996) relata ainda que a “demora desta teoria em ser amplamente utilizada em psicometria consiste na enorme complexidade de manipulação de seus modelos matemáticos” (PASQUALI, 1996, p. 173), tendo recebido maior destaque a partir dos anos 80, com a evolução dos computadores e sistemas informatizados. Por outro lado, Vey (2011) expõe que as bases conceituais da TRI já haviam sido trabalhadas em 1925 por Thurstone, no artigo “*A Method of Scaling Psychological and Educational Tests*”, em que o autor “apresenta uma técnica para a colocação dos itens de uma prova de desenvolvimento mental elaborada por Binet e Simon (1905), em uma escala de classificação por idade” (VEY, 2011, p. 100). Mesmo entendimento é partilhado por Araújo, Andrade e Bortolotti (2009), ao afirmarem que os trabalhos de Thurstone teriam proporcionado grande contribuição para a construção de medidas de traço latente, e na medida de atitude. Expõem os autores:

Em seu trabalho, o autor desenvolveu um método de medida estatístico denominado Lei dos Julgamentos Comparativos que pode ser visto como o mais importante precursor probabilístico da Teoria da Resposta ao Item. No desenvolvimento desse método, Thurstone introduziu dois mecanismos de respostas ou princípios para construção de escalas psicológicas que hoje são conhecidos como mecanismos acumulativo e de desdobramento (ARAÚJO, ANDRADE e BORTOLOTTI, 2009, p. 1001).

Neste cenário, destacam Tezza e Bornia (2009), a Teoria da Resposta ao Item surge para construir escalas padronizadas que independem da amostra e do escore bruto, ou seja, uma forma alternativa de se construir e analisar instrumentos de medida, visto que a Teoria Clássica da Medida “restringe sua validade ao contexto em que foi elaborado, ou seja, os resultados dependem da amostra e as conclusões são baseadas no escore total do teste” (TEZZA e BORNIA, 2009, p. 2).

No Brasil, a TRI tem sido utilizada como técnica para correção das provas do Exame Nacional do Ensino Médio (ENEM). Além da aplicação em testes de conhecimentos e avaliações psicológicas, a TRI vem sendo também aplicada em outras áreas, tais como: gestão da qualidade (ALEXANDRE et al., 2001), avaliação de serviços (BALBIM JUNIOR, 2010), capital humano (VARGAS et al., 2008), avaliação de sites de comércio eletrônico (MOREIRA JUNIOR et al., 2013), e mesmo na área médica (CASTRO et al., 2015).

2.5.2 Conceitos básicos

A Teoria da Resposta ao Item surgiu como uma abordagem alternativa e complementar à Teoria Clássica dos Testes (TCT), como já mencionado, ao possibilitar o desenvolvimento de escalas padronizadas que independem da amostra e do escore bruto. Nesse contexto, Tavares (2001) salienta que a TRI considera cada item particularmente, sem relevar os escores totais. Dessa forma, “as conclusões não dependem propriamente do teste, mas de cada item que o compõe” (TAVARES, 2001, p. 2).

Um paralelo bastante esclarecedor é traçado por Pasquali (2013) ao comparar a Teoria Clássica dos Testes à Teoria da Resposta ao Item:

A Psicometria tradicional (TCT) vai definir a qualidade dos testes psicológicos (...) em termos de um critério, sendo este representado também por comportamentos, a saber, comportamentos presentes ou futuros. (...) Por outro lado, a TRI define a qualidade dos testes (que são comportamentos ou variáveis observáveis) em função de um critério que não é o comportamento e sim variáveis hipotéticas, as quais chama de *teta* ou traço latente (θ) (PASQUALI, 2013, p. 54).

Uma primeira e interessante conceituação à Teoria da Resposta ao Item é dada por Andrade, Tavares e Valle (2000), que enunciam:

A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item (ANDRADE, TAVARES e VALLE, 2000, p. 17).

Esta primeira definição remete principalmente a avaliações de conhecimento ou proficiência, ao contrapor a habilidade do indivíduo e a probabilidade de determinada resposta a um item em um conjunto de testes, o que corrobora o raciocínio de Pasquali (1996). Em seus estudos, Vey (2011) salienta uma das principais finalidades da TRI, ao afirmar:

A proposta da TRI é a de apresentar modelos probabilísticos para variáveis que não são medidas diretamente, tendo como característica principal o item, podendo se entender por item, tarefas ou ações empíricas que constituem a representação do traço latente, ou seja, a habilidade que se pretende medir (VEY, 2011, p. 101).

Antes de avançar no tema, é pertinente a reflexão acerca do conceito de traço latente. Em uma de suas obras sobre psicometria, Pasquali (2013) aborda a falta de consenso sobre a definição de traço latente, tendo por base o ponto de vista de estatísticos e de psicólogos principalmente. Ao afirmar que “traço latente pode ser entendido como um simples parâmetro de caráter matemático e estatístico, como parece entendê-lo a Psicometria Moderna da Teoria da Resposta ao Item” (PASQUALI, 2013, p. 57), o autor o faz com o viés estatístico, que é mais representativo para os fins deste trabalho. Sendo assim, traço latente refere-se a traços hipotéticos não diretamente observáveis – aptidões – que são responsáveis pelos comportamentos emitidos pelos sujeitos – as variáveis observáveis (PASQUALI e PRIMI, 2003; VEY, 2011). Pasquali e Primi (2003) abordam esta conceituação de maneira mais detalhada, ao enunciarem:

O termo teoria do traço latente se refere a uma família de modelos matemáticos que relaciona variáveis observáveis (itens de um teste, por exemplo) e traços hipotéticos não observáveis ou aptidões, estes responsáveis pelo aparecimento das variáveis observáveis ou, melhor, das respostas ou comportamentos emitidos pelo sujeito que são as variáveis observáveis. Assim, temos um estímulo (item) que é apresentado ao sujeito e este responde a ele. A resposta que o sujeito dá ao item depende do nível que o sujeito possui no traço latente ou aptidão. Desta forma, o traço latente é a causa e a resposta do sujeito é o efeito (PASQUALI e PRIMI, 2003, p. 102).

Nojosa (2001, apud VEY, 2011), ao abordar o conceito de traço latente associado à TRI, aponta que esta se apoia em dois pilares:

1 – O desempenho de um indivíduo em um teste pode ser explicado por um conjunto de fatores chamados traços latentes ou habilidades; e

2 – O relacionamento entre as respostas dos indivíduos a cada item e a habilidade medida pelo teste pode ser representada por uma função monótona crescente, chamada Função Característica do Item (FCI). Esta função fornece a probabilidade de indivíduos de vários níveis de habilidade darem certa resposta a um determinado item.

2.5.3 Modelos

A TRI possui vários modelos matemáticos que serão utilizados dependendo principalmente de três fatores, como apontam Andrade, Tavares e Valle (2000): 1) da

natureza do item – dicotômicos ou não dicotômicos (politômicos); 2) do número de populações envolvidas – uma ou mais; e 3) da quantidade de traços latentes a ser medida – apenas um ou mais de um.

Os itens são considerados dicotômicos quando só há a possibilidade de duas respostas. Um item será dicotômico se houver somente duas respostas possíveis, como certo ou errado, no caso de correção de testes, ou quando houver a transformação de determinada escala para apenas duas alternativas – processo denominado dicotomização. No Quadro 1 estão sintetizados os modelos unidimensionais que envolvem uma única população, em relação à natureza do item.

Quadro 1 – Modelos matemáticos para uma única população

Natureza do item	Modelos
Dicotômicos	<ul style="list-style-type: none"> • Logístico de 3 parâmetros (ML3) • Logístico de 2 parâmetros (ML2) • Logístico de 1 parâmetro (ML1)
Não dicotômicos (politômicos)	<ul style="list-style-type: none"> • Modelo de resposta nominal (<i>Nominal response model</i>) • Modelo de resposta gradual (<i>Graded response model</i>) • Modelo de resposta gradual modificado (<i>Modified graded response model</i>) • Modelo de escala gradual (<i>Rating scale model</i>) • Modelo de crédito parcial (<i>Partial credit model</i>) • Modelo de crédito parcial generalizado (<i>Generalized partial credit model</i>) • Modelo de degraus para análise de crédito parcial (<i>Steps model to analyze partial credit</i>) • Modelo de resposta contínua (<i>Continuous response model</i>)

Fonte: adaptado de Vey (2011).

Para facilitar a compreensão da teoria, utilizar-se-á como exemplo a equação básica do Modelo Logístico de 2 Parâmetros (ML2), adotado metodologicamente neste trabalho (Equação 1).

$$P\left(U_{ij} = \frac{1}{\theta_j}\right) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (\text{Equação 1})$$

Onde:

- U_{ij} = variável dicotômica que pode assumir valor 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente;
- θ_j = habilidade ou traço latente do indivíduo j ;
- $P(U_{ij} = 1/\theta_j)$ = probabilidade do indivíduo j com habilidade θ_j responder corretamente o item i ;
- b_i = parâmetro de dificuldade do item i ;
- a_i = parâmetro de discriminação do item i ;
- D = fator de escala, constante e igual a 1.

Vey (2011) assinala que o modelo representa a probabilidade de um indivíduo “ j ” dar uma resposta certa a um item “ i ” de um instrumento de medição (conjunto de itens ou testes) em função dos parâmetros dos itens (a_i , b_i) e do conhecimento θ_j (habilidade, proficiência). Em suma, quanto maior o traço latente, maior a probabilidade de resposta correta.

Em relação aos três modelos propostos para itens de natureza dicotômica, Andrade, Tavares e Valle (2000) apontam serem os modelos de resposta ao item mais utilizados, ao ponto que se diferenciam basicamente pelo número de parâmetros envolvidos, conforme pode ser observado no Quadro 2.

Quadro 2 – Modelos logísticos para itens dicotômicos e parâmetros componentes

Modelo	Parâmetros utilizados
Logístico de 1 parâmetro (ML1)	<ul style="list-style-type: none"> • Dificuldade (b_i)
Logístico de 2 parâmetros (ML2)	<ul style="list-style-type: none"> • Discriminação (a_i) • Dificuldade (b_i)
Logístico de 3 parâmetros (ML3)	<ul style="list-style-type: none"> • Discriminação (a_i) • Dificuldade (b_i) • Probabilidade de acerto ao acaso (c_i)

Adaptado de Andrade, Tavares e Valle (2000).

Os modelos existentes para itens politômicos dependem da natureza das categorias de resposta (VEY, 2011). Estes modelos são adequados para análise de itens com respostas livres ou então de múltipla escolha. Salientam Andrade, Tavares e Valle (2000) que nestes modelos leva-se em consideração não apenas a resposta certa ou errada, mas inclusive qual a resposta dada pelo respondente. O Quadro 3,

adaptado do trabalho de Vey (2011), resume as características dos modelos para itens não dicotômicos.

Quadro 3 – Modelos para itens não dicotômicos e suas características

População	Modelo	Autor(es)	Características
Única	Modelo de resposta nominal	Bock (1972)	Maximiza a precisão do traço latente estimado usando toda a informação contida nas respostas dos indivíduos.
	Modelo de resposta gradual	Samejima (1969)	Assume que as categorias de resposta de um item podem ser ordenadas entre si, podendo-se assim obter mais informação das respostas.
	Modelo de escala gradual	Andrich (1978)	Além das características do modelo de resposta gradual, pressupõe que os escores das categorias sejam igualmente espaçados, como nas escalas tipo Likert, mantendo o mesmo número de categorias de respostas para todos os itens do conjunto.
	Modelo de crédito parcial	Masters (1982)	Extensão do modelo de Rasch, pressupondo que todos os itens tenham o mesmo poder de discriminação e que todos os parâmetros do modelo sejam de localização dos itens. Adequado para usar com qualquer formato de teste que forneça um conjunto de opções de resposta ordenadas.
	Modelo de crédito parcial generalizado	Muraki (1992)	Baseado no modelo de crédito parcial, porém desconsidera a hipótese de uniformidade no poder de discriminação para todos os itens.
Duas ou mais	Modelos adaptados e variações dos já apresentados	Bock e Zimowski (1997)	Generalização dos modelos logísticos de 1, 2 e 3 parâmetros. Possibilita a comparação de indivíduos de grupos distintos, submetidos a conjuntos de itens diferentes, desde que haja itens comuns fazendo a ligação entre os conjuntos de itens.

Fonte: adaptado de Vey (2011, p. 107-108).

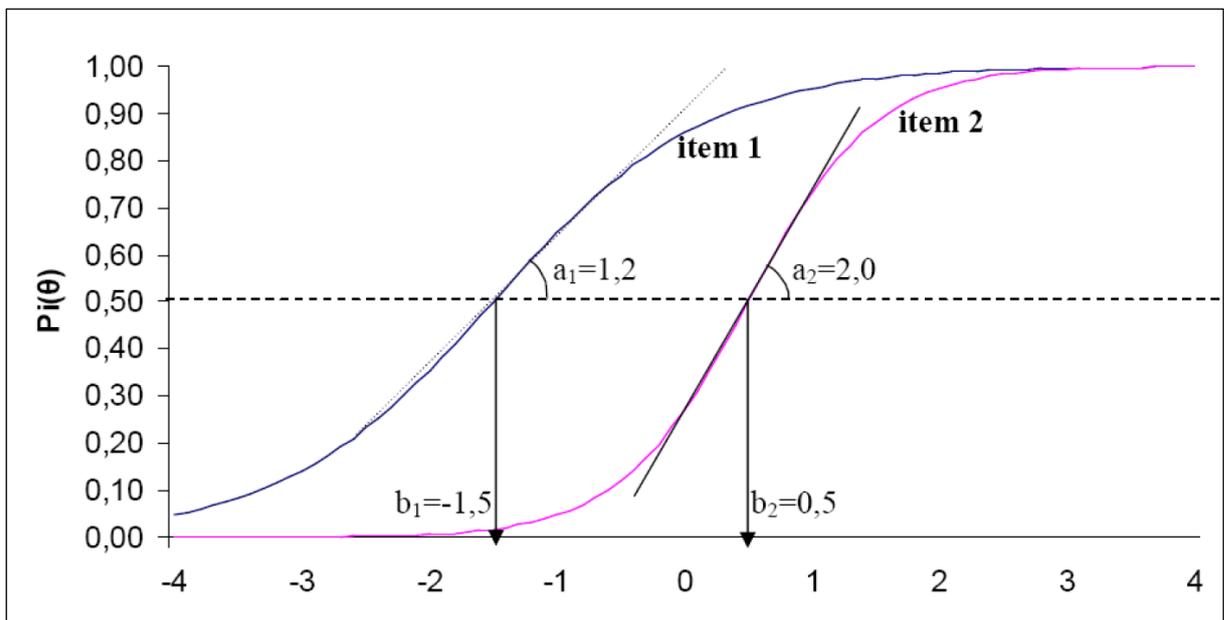
A partir do exposto, verifica-se que a TRI pode ser aplicada em uma grande variedade de estudos. Sua utilização e a seleção do modelo mais adequado para o tipo de pesquisa irá depender do tipo de construto que se deseja medir. Considerando que este estudo avaliou um instrumento que visa medir um único construto, o desempenho docente na ótica dos discentes, optou-se pela utilização de um modelo unidimensional. Além disso, levando-se em conta também que a pesquisa impossibilita o acerto ao acaso, por não se tratar de um teste de conhecimentos, optou-se pela utilização do Modelo Logístico de 2 Parâmetros (ML2).

2.5.3.1 O Modelo Logístico de 2 Parâmetros (ML2)

O Modelo Logístico de 2 Parâmetros é uma variante do ML3, ao suprimir do modelo matemático a variável de acerto ao acaso (c_i), conforme mencionado anteriormente.

Uma das características da TRI enquanto técnica estatística é que o parâmetro de dificuldade do item (b_i) encontra-se na mesma escala do traço latente do indivíduo (θ), permitindo que “a dificuldade dos itens possa ser interpretada em termos de variações padronizadas na habilidade dos sujeitos” (GARCIA et al., 2001 apud COUTO e PRIMI, 2011). O parâmetro a_i (discriminação do item i) é descrito por Tezza e Bornia (2009) como variável que define qual o poder que um item tem de diferenciar indivíduos ou organizações com habilidades semelhantes. A Figura 2 apresenta um exemplo de curva característica para dois itens hipotéticos, e auxilia a interpretação dos parâmetros citados.

Figura 2 – Exemplo de Curva Característica do Item (CCI)



Fonte: Adaptado de Pasquali e Primi (2003).

Analisando-se a figura do ponto de vista do presente estudo, o eixo das abscissas representa a dificuldade dos itens e encontra-se na mesma escala de avaliação do desempenho docente. Sob esse prisma, quanto maior a dificuldade,

maior o grau de avaliação positiva do desempenho (θ) necessário para aquele item. Em outros termos, o parâmetro b_i representa matematicamente a dificuldade do item i , e é definido como o valor em que a probabilidade de resposta certa ao item é de 0,5. A título de exemplificação e trazendo para o contexto deste estudo, discentes posicionados mais à direita na escala (eixo x) serão os que melhor avaliam o desempenho docente, ou seja, apresentam um traço latente maior do que os outros posicionados abaixo na escala. Outra análise possível é que, para itens que possuem b_i maior, a aprovação do desempenho docente dependerá de uma percepção de desempenho alta por parte do respondente (discente). Em sentido oposto, itens que apresentarem um menor valor de b_i serão considerados de mais fácil aprovação de desempenho por parte do respondente.

No que diz respeito ao parâmetro a_i , enuncia Vey (2011) que o parâmetro representa a discriminação do item, e é proporcional à derivada da tangente da curva em seu ponto de inflexão. A interpretação do parâmetro se dá no sentido de que, quanto maior for o valor de a_i , mais inclinada será a curva e, portanto, mais estreito será o intervalo de discriminação do item i .

Retornando ao exemplo da Figura 2, depreende-se que o item 2 é mais difícil do que o item 1. Portanto, apenas discentes com alta percepção do desempenho docente irão aprovar seu desempenho no quesito. A partir disso, pode-se afirmar que os acadêmicos posicionados mais à direita na escala serão aqueles que estarão mais satisfeitos com o desempenho docente, considerando-o elevado e, portanto, aprovando seu desempenho.

2.5.4 Função de informação do item (FII)

A função de informação do item é uma medida bastante utilizada na TRI em conjunto com as curvas características dos itens, pois permite analisar quanto um item contém de informação para a medida de uma habilidade (ANDRADE, TAVARES e VALLE, 2000). Na concepção de Pasquali (2013), a curva de informação do item permite avaliar “o nível de θ para o qual este item traz a maior informação” (PASQUALI, 2013, p. 142). No entendimento de Vey (2011), “a FII reflete a qualidade dos itens individuais e do conjunto de itens como um todo” (VEY, 2011, p. 116), empregado, no estudo em questão, na estimação do desempenho docente e dos

parâmetros dos itens. A função de informação do item no Modelo Logístico de 2 Parâmetros é dada por:

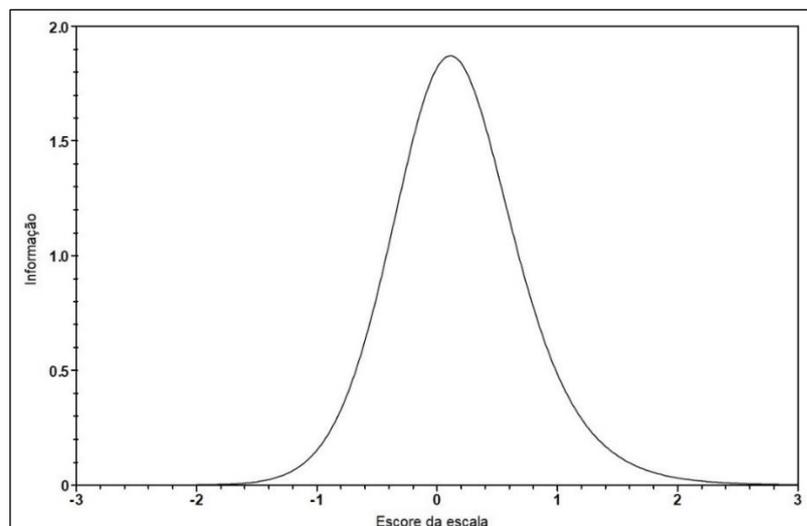
$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta) \quad (\text{Equação 2})$$

Onde:

- $I_i(\theta)$ = informação fornecida pelo item i ao nível de habilidade de θ ;
- a_i = parâmetro de discriminação do item;
- θ = nível de habilidade;
- $P_i(\theta) = 1/(1 + \text{EXP}(-a_i(\theta - b_i)))$;
- $Q_i(\theta) = 1 - P_i(\theta)$.

Com relação à equação, cabe destacar a forte influência do parâmetro a_i , elevado ao quadrado. Partchev (2004 apud VEY, 2011) observa que parâmetros menores do que 1 (um) podem diminuir a FII drasticamente, enquanto parâmetros superiores a 1 (um) agem no sentido contrário, ou seja, tendem a aumentar a função de informação substancialmente. A análise gráfica da Figura 3, onde é apresentado um modelo de curva de informação de um item, facilita a compreensão do conceito.

Figura 3 – Curva de informação de um item



Fonte: adaptado de Vey (2011).

A amplitude da informação é indicada no eixo y, e itens com maior poder de discriminação apresentam uma função de informação mais precisa. Em outras

palavras, uma alta discriminação indica que um item pode diferenciar melhor dois sujeitos que se situam próximos ao valor da dificuldade.

2.5.5 Função de informação do teste (FIT)

No entendimento de Baker (2001 apud VEY, 2011), após um teste ser usado para estimar a capacidade de um examinando, pode-se obter a quantidade de informações geradas pelo teste em qualquer nível de habilidade. Ao considerar um teste como um conjunto de itens, tem-se que a informação do teste em determinado nível de habilidade é a soma das informações do item nesse nível. Dessa forma, pode-se definir a função de informação do teste (FIT) como:

$$I(\theta) = \sum_{i=1}^N I_i(\theta) \quad (\text{Equação 3})$$

Onde:

- $I(\theta)$ = quantidade de informações de um teste a um nível de habilidade (θ);
- $I_i(\theta)$ = quantidade de informações do item i no nível de habilidade (θ);
- N = número de itens do teste.

O entendimento de Andrade, Tavares e Valle (2000) complementa o exposto, pois os autores afirmam que a informação fornecida pelo teste nada mais é do que a soma das informações fornecidas por cada item que o compõe. Os autores apontam ainda que outra maneira de representar a FIT é através do erro-padrão de medida, chamado na TRI de erro-padrão de estimação, dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (\text{Equação 4})$$

Onde:

- $EP(\theta)$ = erro padrão;
- $I(\theta)$ = quantidade de informações de um teste a um nível de habilidade (θ).

Alguns apontamentos importantes são dados por Francisco (2005), ao afirmar que “quanto maior for a função de informação $I(\theta)$, menor será o erro padrão de estimação (EP), e, portanto, maior será a precisão de estimação de habilidade”

(FRANCISCO, 2005, p. 28). O autor assinala ainda que, quanto maior o erro, menor a precisão, fazendo com que seu tamanho dependa de fatores como o número de itens aplicados, da capacidade discriminatória dos itens e da diferença entre o parâmetro b e θ . Nesse sentido, quanto menor a distância entre b e θ , menor será o erro padrão.

2.5.6 Unidimensionalidade e independência local

Uma etapa importante da TRI é a determinação da dimensionalidade do traço latente, ou seja, a quantidade de traços latentes que o instrumento está medindo (VEY, 2011). Segundo Pasquali (2013), um dos postulados da TRI é que há apenas uma aptidão responsável pela realização de um conjunto de tarefas. Apesar de a Psicologia atestar que o desempenho humano é motivado por mais de um traço latente, na TRI considera-se que “é suficiente admitir que haja uma aptidão dominante (um fator determinante) responsável pelo conjunto de itens” (PASQUALI, 2013, p. 84). Couto e Primi (2011), corroboram esse pensamento ao afirmar que a regra adotada na psicometria é a existência de uma dimensão predominante dentre as várias que podem influenciar um comportamento.

Para testar a dimensionalidade, autores como Vey (2011) e Pasquali (2013) sugerem a utilização da análise fatorial de informação plena (*Full Information Factor Analysis – FIFA*) como teste mais indicado para verificar se o construto se mostra unidimensional, podendo-se utilizar o modelo disponível no *software TESTFACT* (VEY, 2011).

A independência local, conforme Araújo, Andrade e Bortolotti (2009), “é obtida quando, controlados pelo nível do traço latente, os itens do teste são independentes, assim a probabilidade de responder um item é precisamente determinada pelo nível do traço latente e não por suas respostas a outros itens do conjunto” (ARAÚJO, ANDRADE e BORTOLOTTI, 2009, p. 1002). Pasquali (2013) traduz muito bem essa afirmação ao colocar que, mantidas constantes as aptidões que afetam o teste, as respostas dos indivíduos a quaisquer dos itens serão estatisticamente independentes. Dessa forma, a independência local é tida como consequência da correta determinação da dimensionalidade dos dados (LORD, 1980, apud BORTOLOTTI et al., 2012).

2.5.7 Estimação

A estimação dos parâmetros dos itens e as proficiências dos respondentes nos modelos logísticos da TRI é considerada por Andrade, Tavares e Valle (2000) como uma das etapas mais importantes de sua aplicação. Nesse sentido, Vey (2011) reforça que a probabilidade de um sujeito dar uma resposta positiva a um determinado item depende de sua habilidade, e dos parâmetros que caracterizam o item, e que estes, geralmente são desconhecidos. Na fase de estimação busca-se, portanto, determinar o valor de θ_j para cada indivíduo e os parâmetros (a_i e b_i) de cada item do teste.

Couto e Primi (2011) apresentam em seu trabalho um conjunto de passos a serem desenvolvidos para se chegar a estimação dos itens de um teste. Nele, os autores afirmam que o primeiro passo é comum ao processo de construção de instrumentos de medida de maneira geral, e consiste na “definição rigorosa da dimensão que se pretende avaliar, seguida da elaboração de itens que representem adequadamente essa dimensão” (COUTO e PRIMI, 2011, p. 7).

O segundo passo consiste na aplicação dos itens a uma amostra representativa da população para a qual se está construindo o instrumento, inicialmente com o intuito de estimar os parâmetros com base na teoria clássica e verificar a unidimensionalidade dos itens. Conforme estudo de Nunes e Primi (2005), amostras a partir de 200 respondentes são consideradas significativas para a boa estimação dos parâmetros. Araújo, Andrade e Bortolotti (2009) apontam que existem diversos métodos de estimação, sendo que os mais usados são o Método da Máxima Verossimilhança e os Métodos Bayesianos.

Em seu trabalho, Vey (2011) indica que as equações de estimação utilizadas por esses métodos não podem ser solucionadas analiticamente, e necessitam, portanto, serem resolvidas com o uso de procedimentos numéricos iterativos, como o algoritmo Newton-Raphson e o método *scoring* de Fischer. Andrade, Tavares e Valle (2000) relatam que, na estimação dos parâmetros dos itens, usualmente chamada calibração, é mais frequente a aplicação do método da Máxima Verossimilhança Marginal. Já para estimação dos traços latentes, é mais comum a aplicação do método Bayesiano.

O Quadro 4 apresenta os principais métodos de estimação disponíveis, seus autores e uma descrição resumida de cada um.

Quadro 4 – Métodos de estimação e seus autores

Método	Autor	Descrição
Máxima Verossimilhança Conjunta	Lord (1974, 1980)	Os parâmetros dos itens e da performance são estimados simultaneamente.
Máxima Verossimilhança Marginal	Bock e Aitkin (1981)	A estimação é realizada em duas etapas. Os parâmetros das performances são integrados em separado e os parâmetros dos itens são estimados.
Máxima Verossimilhança Condicional	Andersen (1972, 1973); Rasch (1960)	A função verossimilhança está condicionada ao número de escores brutos.
Estimação Bayesiana Conjunta e Marginal	Mislevy (1986); Swaminathan e Gifford (1982, 1985, 1986)	Estabelece distribuições a priori para os parâmetros dos itens e das performances eliminando alguns problemas, como estimação de parâmetros imprópria e não convergência, encontrados com os métodos de Máxima Verossimilhança conjunta e marginal.

Fonte: extraído de Vargas (2007).

No presente estudo, em que a proposta é avaliar o instrumento de medida, não foram estimadas as proficiências dos respondentes, apenas os parâmetros dos itens a partir do método da Máxima Verossimilhança Marginal.

2.5.8 Vantagens e limitações da TRI

A Teoria da Resposta ao Item possui um campo de aplicação bastante amplo, conforme demonstram estudos encontrados nas áreas de gestão da qualidade (ALEXANDRE et al., 2001), avaliação de serviços (BALBIM JUNIOR, 2010), capital humano (VARGAS et al., 2008), entre outros. O número de publicações na área tem contribuído para sua divulgação. Salienta Vey (2011) que a TRI não veio para substituir a Teoria Clássica dos Testes, mas para complementá-la na medida em que permite a análise individual dos itens e a fidedignidade da medida. Além disso, o crescente número de publicações em variadas áreas deriva das particularidades dos modelos disponíveis.

Algumas das vantagens da TRI sobre a Teoria Clássica dos Testes e a Teoria Clássica de Medidas são apontadas por Araújo, Andrade e Bortolotti (2009), e elencadas no trabalho de Mello (2014):

- Possibilita que sejam feitas comparações entre traço latente e indivíduos de populações diferentes que tenham sido submetidos a questionários que tenham alguns itens em comum;

- Permite que indivíduos de uma mesma população sejam comparados mesmo que submetidos a testes totalmente diferentes, tendo em vista que o elemento central da TRI é o item e não o teste como um todo;
- Possibilita uma melhor análise de cada item do instrumento de medida, uma vez que considera suas características específicas de construção de escalas;
- Os itens e o nível de conhecimento dos indivíduos encontram-se na mesma escala, permitindo a comparação entre o nível do respondente e o nível da característica requerida pelo item. Isto é particularmente importante pois facilita a interpretação da escala criada, e identificar quais itens estão produzindo informação ao longo da mesma;
- Permite o tratamento para um conjunto de dados faltantes, ao utilizar somente os dados respondidos;
- Princípio da invariância, ou seja, os parâmetros dos itens independem do traço latente do respondente e os parâmetros dos indivíduos não dependem dos itens apresentados.

Apesar dos avanços em relação às teorias clássicas, a TRI também apresenta limitações. A principal delas envolve o tamanho da amostra necessária para estimação dos parâmetros. Bortolotti et al. (2012) apregoam que, para se obter estimativas dos parâmetros dos itens com baixos erros padrão, faz-se necessária uma amostra significativa de respondentes. Além disso, o número de respondentes depende do número de parâmetros a serem estimados e do número de categorias de respostas. Outro aspecto apontado por Vey (2011) em relação a uma amostra ideal, é que se tenham respondentes nos diversos níveis da escala, o que gera uma preocupação não só com o tamanho da amostra, mas também com sua qualidade.

3 METODOLOGIA

A partir do momento em que foram definidos o problema de pesquisa e os objetivos, o desenvolvimento do método deve ser pautado com a finalidade de propor respostas adequadas e, dentro do possível, com maior nível de acuidade. Método é definido por Cervo, Bervian e Silva (2007, p. 27) como “a ordem que se deve impor aos diferentes processos necessários para atingir um certo fim ou um resultado desejado”. Ainda nas palavras dos autores, considerando o viés das Ciências Sociais, o método é constituído pelo conjunto de processos utilizados na investigação e demonstração da verdade, ou seja, para que se obtenha uma ou mais respostas ao problema de pesquisa. Pacheco Júnior et al. (2007, apud VEY, 2011), relata que é comum a confusão entre os conceitos de método e metodologia. Para o autor, método refere-se à maneira utilizada para se atingir determinado objetivo, enquanto metodologia refere-se ao estudo do método (VEY, 2011).

3.1 CARACTERIZAÇÃO DA PESQUISA

O presente trabalho adotou classificação semelhante àquela utilizada por Vey (2011), em razão das similaridades entre esta dissertação e a tese elaborada pelo autor, guardadas as devidas proporções, no que diz respeito à utilização da Teoria da Resposta ao Item. Complementarmente, foi utilizada a classificação proposta por Gil (2010) para delinear a pesquisa e estruturar o trabalho.

3.1.1 Base filosófica

As principais correntes filosóficas em Ciências Sociais que abordam a validade do conhecimento são o positivismo, o estruturalismo e o materialismo dialético (MELLO, 2014). Nesse sentido, Vey (2011) apresenta uma reflexão no que tange a estas correntes. Para o autor, “uma pesquisa deve seguir uma perspectiva epistemológica a qual orientará o pesquisador na escolha do método, metodologia e técnicas que irá utilizar na sua pesquisa” (VEY, 2011, p. 147). Considerando este enfoque, este trabalho foi centrado na base filosófica estruturalista. O entendimento de Motta (1970) sobre estruturalismo, apesar ter sido exposto há mais de 40 anos, é bastante elucidativo quanto ao conceito:

A palavra estrutura é de emprego muito antigo, tanto nas ciências físicas quanto nas sociais e em termos amplos significa tudo o que a análise interna de uma totalidade revela, ou seja, elementos, suas inter-relações, disposição. O conceito de estrutura é especialmente importante para a ciência porque pode ser aplicado a coisas diferentes, permitindo a comparação. Nesse sentido, podemos afirmar que o estruturalismo é um método analítico comparativo. Além disso, o estruturalismo considera os fenômenos ou elementos com referência a uma totalidade, considerando, pois, o seu valor de posição (MOTTA, 1970, p. 24).

O elemento chave do estruturalismo está centrado no estudo das inter-relações que explicam um fenômeno (VEY, 2011).

3.1.2 Classificações de pesquisa

Gil (2010) propõe uma série de classificações em que um trabalho científico pode ser enquadrado, segundo diversos fatores. O autor destaca a importância da classificação por facilitar a organização dos fatos e seu entendimento (GIL, 2010).

Inicialmente, faz-se necessária a classificação conforme recomendam Marconi e Lakatos (2003), no tocante às técnicas de abordagem, pensamento ou raciocínio. De acordo com esta categorização, o método de pesquisa pode ser indutivo, dedutivo, hipotético-dedutivo ou ainda, dialético. Neste quesito, o método mais adequado para atingimento dos objetivos da pesquisa e para resolução do problema em questão é o *dedutivo*. No entendimento de Cerro, Bervian e Silva (2007, p. 46), a dedução “é a argumentação que torna explícitas verdades particulares contidas em verdades universais. O ponto de partida é o antecedente, que afirma uma verdade universal, e o ponto de chegada é o conseqüente, que afirma uma verdade particular”.

Quanto à sua natureza, no entendimento de Silva e Menezes (2000), as pesquisas podem ser classificadas como básicas ou aplicadas. A presente pesquisa classifica-se como *aplicada*, pois tem como finalidade “a aquisição de conhecimento com vistas à aplicação numa situação específica” (GIL, 2010, p. 27).

Em relação à forma de abordagem do problema, a pesquisa pode ser caracterizada como qualitativa ou quantitativa (SILVA e MENEZES, 2010). Esta pesquisa tem caráter predominantemente *quantitativo*, dado o enfoque estatístico presente na Teoria da Resposta ao Item. Além disso, o instrumento de pesquisa a ser analisado é composto por questões eminentemente quantitativas, baseadas em uma escala tipo Lickert. Nas palavras de Richardson (2010, apud VEY, 2011), o método

quantitativo é utilizado em larga escala em estudos descritivos que procuram analisar a relação entre variáveis, bem como para investigar a relação de causalidade entre fenômenos.

No que diz respeito aos objetivos da pesquisa, Silva e Menezes (2000) propõem três categorias: exploratória, descritiva ou explicativa. O presente estudo apresenta características de pesquisas descritivas e exploratórias. Salientam Cervo, Bervian e Silva (2007):

(...) a pesquisa descritiva observa, registra, analisa e correlaciona fatos ou fenômenos (variáveis) sem manipulá-los. Procura descobrir, com a maior precisão possível, a frequência com que um fenômeno ocorre, sua relação e conexão com outros, sua natureza e suas características (CERVO, BERVIAN e SILVA, 2007, p. 61).

O caráter exploratório, destacam Silva e Menezes (2000), surge pois busca-se a familiarização com o problema, para torná-lo explícito ou para construção de hipóteses.

3.1.3 Procedimentos metodológicos e técnicas de pesquisa

A classificação quanto a procedimentos técnicos e metodológicos empregados proposta por Gil (2010) inclui ao menos treze delineamentos distintos, a saber: 1) pesquisa bibliográfica; 2) pesquisa documental; 3) pesquisa experimental; 4) ensaio clínico; 5) estudo caso-controle; 6) estudo de coorte; 7) levantamento de campo (*survey*); 8) estudo de caso; 9) pesquisa etnográfica; 10) pesquisa fenomenológica; 11) teoria fundamentada nos dados (*grounded theory*); 12) pesquisa-ação; e 13) pesquisa participante.

O presente estudo centrou-se, principalmente, na pesquisa bibliográfica para construção do corpo teórico, e na pesquisa documental, uma vez que a base de dados para tratamento e análises estatísticas foi disponibilizada pelo Centro de Processamento de Dados (CPD) da UFSM. Gil (2010), ao diferenciar pesquisa bibliográfica de documental principalmente no que concerne a fonte dos dados – a primeira sendo fundamentada em material elaborado por autores com propósito de ser lido por públicos específicos, e a segunda elaborada com finalidades diversas –, exemplifica alguns casos em que fontes podem ser consideradas ora bibliográficas, ora documentais (relatórios, boletins, atos jurídicos, compilações estatísticas, etc.).

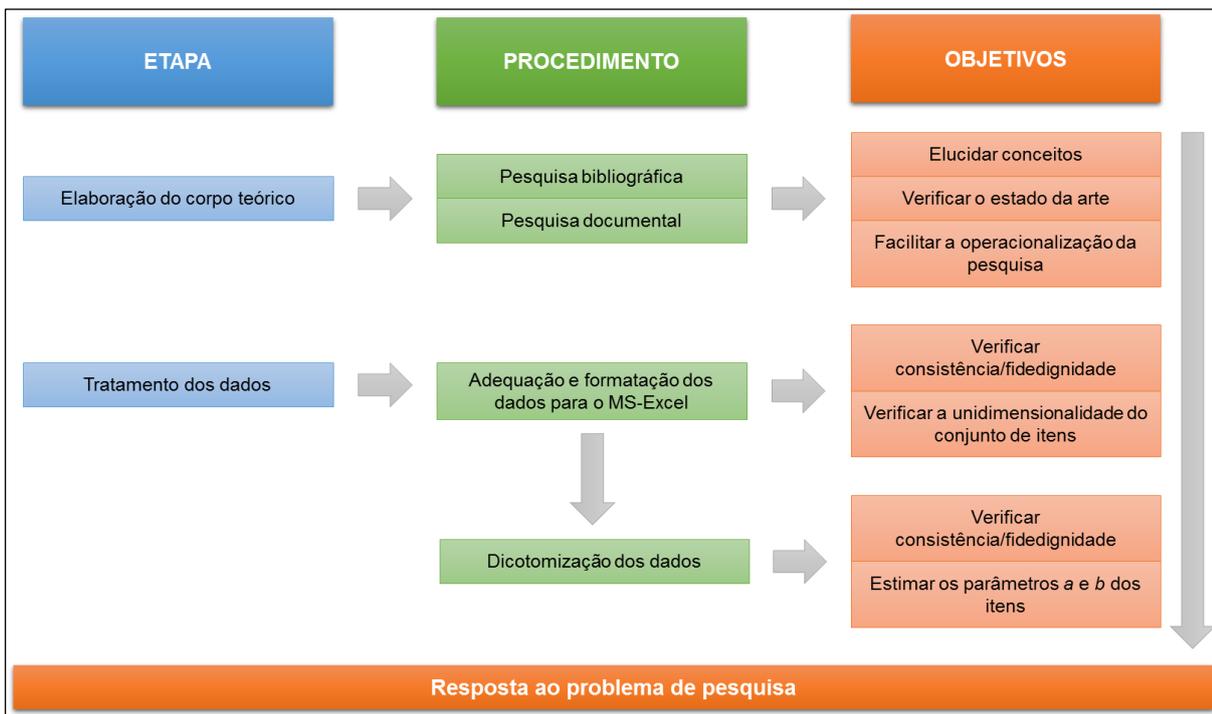
Entretanto recomenda que “seja considerada fonte documental quando o material consultado é interno à organização, e fonte bibliográfica quando for obtido em bibliotecas ou bases de dados” (GIL, 2010, p. 31).

A pesquisa bibliográfica, que possibilitou a elaboração da estrutura teórica do trabalho para lhe dar sustentação, envolveu a consulta a livros, anais de congressos, periódicos nacionais e internacionais, além de dissertações e teses.

3.2 ETAPAS DA PESQUISA

O planejamento da pesquisa e a definição de suas etapas é um elemento crucial que tende a facilitar as fases que sucedem, auxiliando o alcance dos objetivos propostos. As etapas da pesquisa são sintetizadas na Figura 4.

Figura 4 – Etapas da pesquisa



Fonte: elaborado pelo autor.

Comum a maioria dos trabalhos científicos, este teve início com a pesquisa bibliográfica. Com o objetivo elucidar os conceitos centrais referentes ao tema, a formulação dos construtos e a verificação do estado da arte, a pesquisa bibliográfica

“procura explicar um problema a partir de referências teóricas publicadas em artigos, livros, dissertações e teses” (CERVO, BERVIAN e SILVA, 2007, p. 60). Os autores ressaltam que se trata de uma forma de se conhecer e analisar as contribuições científicas já existentes, e geralmente constituem o passo inicial de qualquer trabalho científico. Nesta etapa foram aprofundados conceitos relativos à TRI, à avaliação docente, à avaliação de desempenho e a avaliação do ensino superior no Brasil.

Concluída a primeira etapa, passou-se à parte prática do estudo. Neste momento, foi efetuada a análise do instrumento de pesquisa da avaliação docente, aplicado pela CPA aos discentes dos cursos de graduação presencial do campus Santa Maria da UFSM, em fase de testes, no período entre 16 de junho e 16 de julho de 2016. A participação na avaliação foi voluntária, e as respostas foram coletadas em formulário eletrônico disponibilizado no Portal do Aluno.

A análise foi realizada inicialmente sob a ótica da Teoria Clássica dos Testes (TCT), e em seguida sob a perspectiva da TRI. Para tanto, foi utilizada a base de dados disponibilizada em documento do *MS-Excel* pela Coordenadoria de Planejamento e Avaliação Institucional (COPLAI), contendo as respostas dadas pelos acadêmicos. Após a formatação da base de dados, esta foi exportada para o *software Statistical Package for the Social Sciences (SPSS)*, para que fossem rodados os testes alfa de Cronbach, com o objetivo de avaliar a fidedignidade e a consistência interna do conjunto de itens para o construto (VEY, 2011), e a análise fatorial exploratória pelo método das componentes principais. Conforme recomenda Nunnally (apud SILVA FILHO, 2015), para uma investigação preliminar, um alfa de Cronbach igual ou maior a 0,7 é aceitável.

Executados os testes iniciais, a base de dados foi dicotomizada, formando um conjunto de dados contendo a totalidade das respostas dadas pelos discentes da UFSM, ao que se convencionou chamar dados agrupados ou agregados, e outros dez subconjuntos contendo as avaliações recebidas pelos docentes de cada Centro de Ensino ou Unidade Universitária, a saber: Centro de Artes e Letras (CAL); Centro de Ciências Naturais e Exatas (CCNE); Centro de Ciências Rurais (CCR); Centro de Ciências da Saúde (CCS); Centro de Ciências Sociais e Humanas (CCSH); Centro de Educação (CE); Centro de Educação Física e Desportos (CEFD); Centro de Tecnologia (CT); Colégio Técnico Industrial de Santa Maria (CTISM); e Colégio Politécnico (Politécnico).

Após o processo de dicotomização, novamente realizou-se a análise fatorial, neste caso, a análise fatorial de informação plena, própria para itens dicotomizados, para verificar a unidimensionalidade dos itens para o conjunto agregado de respostas, utilizando o *software TESTFACT*, conforme sinalizam Tavares, Andrade e Valle (2000). A seguir, foram estimados os parâmetros de discriminação e dificuldade dos itens (a e b , respectivamente), utilizando-se o Modelo Logístico de 2 Parâmetros (ML2). Os parâmetros foram calculados através do *software Bilog-MG* pelo método da Máxima Verossimilhança Marginal. Executados os testes, foram apresentadas as curvas características dos itens (CCI), que balizaram as interpretações do estudo.

Por fim, foram apresentadas as considerações finais e as sugestões para pesquisas futuras.

3.3 LIMITAÇÕES DO ESTUDO

As principais limitações do estudo identificadas dizem respeito à caracterização da amostra e à população, o que limitou a profundidade das análises. Em primeiro lugar, a base de dados não permitiu identificar o vínculo dos respondentes (em qual curso o discente estaria matriculado). Esta informação permitiria caracterizar melhor a amostra considerada.

Outra limitação decorre de questões éticas, pois a identificação dos docentes avaliados ou ao menos do departamento no qual ele exerce suas atividades, reduziu a possibilidade de algumas análises baseadas na Teoria Clássica dos Testes. Por outro lado, esta informação poderia causar constrangimento entre os membros da categoria, e dificilmente conseguir-se-ia autorização formal e adesão de um número suficientemente grande de docentes para viabilizar a pesquisa dessa forma.

Além destes fatores, outras limitações podem decorrer das citadas distorções possíveis à avaliação de desempenho, como o efeito halo, erro de tendência central ou brandura/rigor.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise dos resultados obtidos teve como delineamento principal os objetivos específicos do trabalho, com vistas a atingir o objetivo geral e responder ao problema de pesquisa. Assim, foram realizados os seguintes procedimentos, após o tratamento dos dados: teste de alfa de Cronbach; análise fatorial exploratória pelo método de componentes principais; correlação bisserial; análise fatorial de informação plena; e análise dos parâmetros de discriminação e dificuldade dos itens (parâmetros a_i e b_i , respectivamente).

4.1 TRATAMENTO DOS DADOS

Foram analisados os 13 primeiros itens do instrumento de medida aplicado, os quais estão diretamente ligados ao desempenho docente, pois os itens restantes (14, 15, 16 e 17) tratam de questões peculiares e individuais, relativas à docência orientada (14 a 16), e ao comprometimento do aluno com as atividades propostas pelo professor (17).

Inicialmente, os dados foram tabulados e analisados utilizando a Teoria Clássica dos Testes utilizando os programas *MS-Excel* e *SPSS*. Posteriormente, os dados foram dicotomizados e analisados com base na Teoria da Resposta ao Item utilizando os programas *TESTFACT* e *Bilog-MG*, específicos para a TRI.

A avaliação ocorreu no primeiro semestre do ano de 2016. Foram avaliados todos os docentes pertencentes ao quadro da UFSM. Considerando os dados tabulados no *MS-Excel*, em que cada linha preenchida representa a avaliação dada por um acadêmico ao professor de uma disciplina, foram respondidos mais de 23.000 questionários somente no campus sede.

Tendo em vista os objetivos do trabalho e o volume de dados disponíveis, a análise baseada na TCT limitou-se aos resultados gerais, não entrando no âmbito das Unidades Universitárias (Centros de Ensino). A análise através da TRI apresentou maior nível de detalhes, ao focar comparativamente os resultados obtidos pelas Unidades, e relacioná-los ao resultado obtido pelo conjunto agregado de dados, que incluiu a totalidade das respostas ao instrumento de medida. O objetivo disto é verificar se o comportamento dos itens nas Unidades Universitárias se relaciona com o resultado da Instituição.

4.2 ANÁLISE BASEADA NA TEORIA CLÁSSICA DOS TESTES (TCT)

Uma das propriedades na Teoria Clássica dos Testes utilizada para avaliar a qualidade do instrumento de medida é dada pela fidedignidade. A consistência interna do conjunto de itens que medem o desempenho docente pode ser verificada através da determinação do alfa de Cronbach. Para Nunnally (1978), um valor superior a 0,75 é considerado satisfatório. Nesse sentido, Pasquali (2013) coloca que quando o resultado do coeficiente se aproxima de 1 (um), pode-se afirmar que o teste possui um coeficiente de precisão.

Na Tabela 1 são apresentados os dados descritivos da amostra em que este estudo se baseou. Após, são apresentados os resultados do teste alfa de Cronbach.

Tabela 1 – Número de respondentes

	N	%
Válidos	20543	86,1
Excluídos	3328	13,9
Total	23871	100,0

Fonte: elaborado pelo autor com resultados do *software SPSS*.

Foram analisadas 23.871 respostas aos itens, com 20.543 destas sendo consideradas na análise do alfa de Cronbach. Apesar de o número total de casos considerados no teste ser de 23.871, dezessete deles foram excluídos nas demais análises em razão da atual lotação de exercício dos docentes avaliados ser a Reitoria Universitária. Isto impossibilitou que, no banco de dados, fosse identificado o departamento vinculado à Unidade Universitária a qual o docente está oficialmente lotado. A Tabela 2 traz o quantitativo de avaliações por Unidade Universitária e a participação percentual em relação a totalidade de casos.

Tabela 2 – Número de avaliações por Unidade Universitária

(continua)

Unidade Universitária	Avaliações	Participação percentual
Centro de Artes e Letras (CAL)	1.551	6,50%
Centro de Ciências Naturais e Exatas (CCNE)	4.314	18,09%

Tabela 2 – Número de avaliações por Unidade Universitária

(conclusão)

Unidade Universitária	Avaliações	Participação percentual
Centro de Ciências Rurais (CCR)	3.309	13,87%
Centro de Ciências da Saúde (CCS)	2.971	12,45%
Centro de Ciências Sociais e Humanas (CCSH)	4.208	17,64%
Centro de Educação (CE)	1.739	7,29%
Centro de Educação Física e Desportos (CEFD)	506	2,12%
Centro de Tecnologia (CT)	4.247	17,80%
Colégio Técnico Industrial (CTISM)	339	1,42%
Colégio Politécnico	670	2,81%
TOTAL	23.854	100,00%

Fonte: elaborado pelo autor com resultados do *software* SPSS.

Destaca-se na Tabela 2 a participação de três Unidades Universitárias: Centro de Ciências Naturais e Exatas (CCNE), Centro de Tecnologia (CT) e Centro de Ciências Sociais e Humanas (CCSH). O CCNE, apesar de ter 1.410 alunos em cursos de graduação oferecidos pelo Centro de Ensino, apresentou maior número de avaliações em relação aos demais. Isto pode ser reflexo do modelo de estrutura adotado, departamental, tendo em vista que os Departamentos do CCNE atendem também a cursos de graduação ofertados por outros Centros, como é o caso das disciplinas iniciais dos cursos de tecnologia, envolvendo Física, Cálculo, entre outras. Já a participação dos Centros de Tecnologia, e de Ciências Sociais e Humanas, se dá por se tratarem das duas maiores Unidades Universitárias em número de alunos de graduação – 3.066 e 3.516, respectivamente, conforme dados de 2016 disponíveis no *website* “UFSM em Números”.

As Tabelas 3 e 4 apresentam as estatísticas individuais de cada item, e o resultado final do teste alfa de Cronbach para o conjunto, considerado o quantitativo da Tabela 1.

Tabela 3 – Alfa de Cronbach dos itens

(continua)

Item	Alfa de Cronbach
01	0,940
02	0,943
03	0,936
04	0,939
05	0,938

Tabela 3 – Alfa de Cronbach dos itens

(conclusão)

Item	Alfa de Cronbach
06	0,938
07	0,936
08	0,938
09	0,937
10	0,940
11	0,939
12	0,942
13	0,939

Fonte: elaborado pelo autor com resultados do *software SPSS*.

Tabela 4 – Alfa de Cronbach para o conjunto de itens

Alfa de Cronbach	Alfa de Cronbach com base em itens padronizados	N de itens
0,943	0,944	13

Fonte: elaborado pelo autor com resultados do *software SPSS*.

Considerando que o resultado final do alfa de Cronbach para o conjunto de itens foi de 0,94, pode-se afirmar que o instrumento é confiável e possui consistência interna.

Outro procedimento adotado neste estudo foi a utilização da análise fatorial para testar a dimensionalidade do conjunto de itens. A proposta primordial da análise fatorial é descrever, se possível, a estrutura de covariâncias entre as variáveis em termos de um número menor de variáveis (não observáveis) chamadas fatores. Por outras palavras, a análise fatorial estuda os inter-relacionamentos entre as variáveis, num esforço para encontrar um conjunto de fatores (em menor número que o conjunto de variáveis originais) que exprima o que as variáveis originais partilham em comum.

Assim, para testar a dimensionalidade, Spencer (2004) sugere a utilização da análise fatorial que, para o autor, é um conjunto de ferramentas estatísticas utilizadas para reduzir o número de variáveis, assim como para avaliar a estrutura de dados, podendo ser exploratória ou confirmatória. Mello (2014) complementa este entendimento ao afirmar que a técnica se propõe a analisar a estruturas das inter-relações entre um grande número de variáveis, ao passo que define um conjunto de dimensões latentes – os fatores.

Nesta análise utilizou-se o método de componentes principais como método de estimação. Na Tabela 5 são apresentados os resultados extraídos com a utilização do *software SPSS*.

Tabela 5 – Variância total explicada

Componente	Valores próprios iniciais			Somadas de extração de carregamentos ao quadrado		
	Total	% de variância	% cumulativa	Total	% de variância	% cumulativa
1	7,813	60,102	60,102	7,813	60,102	60,102
2	,865	6,654	66,756			
3	,682	5,246	72,003			
4	,580	4,461	76,464			
5	,524	4,034	80,498			
6	,459	3,528	84,026			
7	,396	3,047	87,073			
8	,365	2,805	89,877			
9	,317	2,438	92,315			
10	,279	2,148	94,463			
11	,262	2,019	96,482			
12	,239	1,836	98,318			
13	,219	1,682	100,000			

Fonte: elaborado pelo autor com resultados do *software SPSS*.

Com base nos resultados apresentados pode-se afirmar que existe um fator (variável) determinante responsável por 60,10% da variabilidade dos dados. O segundo fator de maior representatividade responde por 6,65% da variância total. Com base nisso, verifica-se que o primeiro fator é responsável por explicar mais de nove vezes a variação do conjunto de itens do que o segundo fator. Na concepção de Hair Junior et al. (2005), é comum em ciências sociais que seja considerada satisfatória uma solução que explique 60% da variância total. Nesse sentido, estes valores atestam que o instrumento de medida é unidimensional, ou seja, está medindo um único traço latente, o desempenho dos docentes.

Assim, com base nos resultados estatísticos desta primeira análise, a qual testou a confiabilidade e dimensionalidade do instrumento de medida aplicado junto aos discentes, conclui-se que o instrumento de medida possui confiabilidade e consistência interna. Além disso, o conjunto de itens que compõe o instrumento de

medida estão medindo um único traço latente, o desempenho docente, ou seja, o instrumento de medida é unidimensional.

4.3 ANÁLISE BASEADA NA TEORIA DA RESPOSTA AO ITEM (TRI)

4.3.1 Análise do resultado agregado das Unidades Universitárias

Nesta parte da análise o estudo utilizou-se da Teoria da Resposta ao Item. Cada subseção a seguir abordou uma das etapas metodológicas para atendimento dos objetivos propostos pelo trabalho.

4.3.1.1 Dicotomização dos itens

A primeira etapa metodológica para a realização da análise com base na TRI consiste na dicotomização das respostas atribuídas pelos discentes para cada um dos itens. Esta etapa fez-se necessária em razão de o instrumento aplicado ter utilizado uma escala tipo Likert de 5 pontos, conforme exposto na Tabela 6.

Tabela 6 – Escala de concordância

1	2	3	4	5
Discordo totalmente	Discordo em parte	Não concordo nem discordo	Concordo em parte	Concordo totalmente

Fonte: CPA/UFSM.

Após análise da média e mediana dos itens, decidiu-se cortar todos os itens no 3, ou seja, ao discente que respondeu 1, 2 ou 3, considerou-se como “não aprova o desempenho do docente” ($x = 0$) naquele item. Já aos discentes que responderam 4 ou 5, considerou-se que estes “aprovam o desempenho do docente” ($x = 1$).

4.3.1.2 Análise fatorial

Com os dados dicotomizados, os itens novamente foram testados para verificar a dimensionalidade do instrumento de medida. Desta vez, a análise da

dimensionalidade foi realizada através da análise fatorial de informação plena (*Full Information Factor Analysis – FIFIA*), utilizando-se o *software TESTFACT*.

O resultado obtido apresentou o primeiro fator responsável por explicar 63,70% da variabilidade geral dos dados e o segundo fator 4,52%, denotando que o primeiro fator é responsável por explicar mais de quatorze vezes a variação do conjunto de itens do que o segundo fator. Assim, a análise comprova a suposição inicial de que o conjunto de itens é unidimensional, ou seja, está medindo um único traço latente, o desempenho docente. Isso quer dizer que o resultado justifica a utilização de um modelo unidimensional da TRI, no caso, o Modelo Logístico de 2 Parâmetros (VEY, 2011).

4.3.1.3 Correlação bisserial

Na Tabela 7 são apresentados os dados da correlação bisserial. No entendimento de Soares e Mendonça (2003), a correlação bisserial é uma medida estatística da capacidade de discriminação do item, na medida em que verifica se um determinado item binário apresenta correlação significativa com o escore bruto resultante do conjunto de itens.

Tabela 7 – Correlação bisserial do conjunto de itens.

Item	Correlação bisserial
01	0,9420
02	0,8330
03	1,0160
04	0,8350
05	1,0000
06	0,9280
07	1,0750
08	1,0630
09	0,9810
10	0,9780
11	0,9760
12	0,7860
13	1,0290

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

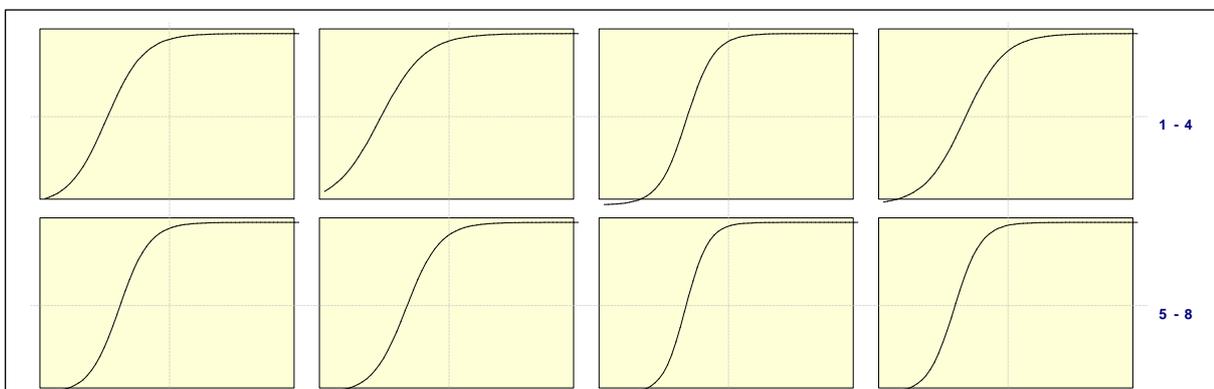
A partir da análise dos dados da correlação bisserial, pode-se afirmar que há forte correlação entre cada item e o conjunto como um todo, tendo em vista que é usual a aceitação de itens que apresentem valores superiores a 0,3 para este parâmetro (SOARES, 2005). Considerando que o item que apresentou menor correlação (item 12) atingiu 0,7860, tem-se um primeiro indício de que o conjunto de itens do instrumento de coleta de dados foi bem construído, ao apresentarem forte consistência interna e correlação com o escore bruto.

4.3.1.4 Estimação dos parâmetros dos itens

A estimação dos parâmetros a e b considera que os dados seguem uma distribuição normal, com média igual a zero e desvio padrão igual a um ($\mu=0$ e $\sigma=1$). Conforme atesta Vey (2011), esta etapa dos testes estatísticos é conhecida na TRI como calibração dos itens. Para tanto, utilizou-se o método de Máxima Verossimilhança Marginal (MVM) na estimação dos parâmetros. Ademais, a convergência dos dados também foi testada pelo algoritmo *Expectation – Maximization* (EM) e Newton-Raphson (NR), no intuito de garantir a boa estimação dos parâmetros (MELLO, 2014). No caso do banco de dados agregado, em que estão inclusas todas as Unidades Universitárias, a convergência do conjunto de itens foi atingida em 14 ciclos no algoritmo EM, e em 16 ciclos através do algoritmo NR.

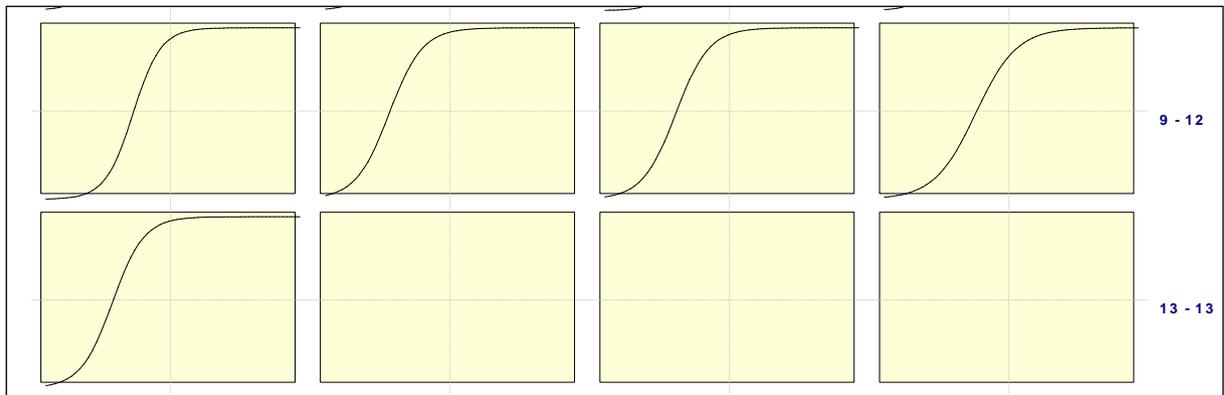
Nas Figuras 5 e 6 são apresentadas as curvas características dos itens (CCI).

Figura 5 – Curvas características dos itens 1 a 8



Fonte: elaborado pelo autor.

Figura 6 – Curvas características dos itens 9 a 13



Fonte: elaborado pelo autor.

As Curvas Características dos itens são analisadas em conjunto com a Tabela 8, em que constam os valores dos parâmetros a e b – discriminação e dificuldade, respectivamente – de cada item integrante do teste, e as assertivas correspondentes aos itens.

A interpretação a partir das CCI do parâmetro a é feita analisando-se o ângulo de inclinação no ponto de inflexão. Nesse ponto, existe 50% de probabilidade de um indivíduo responder corretamente a um item no caso de um teste de conhecimentos ou, em se tratando de avaliação de desempenho, 50% de chances de aprovar o desempenho. Este parâmetro pode ser considerado um indicador de qualidade do item, na medida em que, quanto maior seu valor, maior será sua capacidade de diferenciar indivíduos com habilidades semelhantes.

Em relação ao parâmetro b , pode-se extrair da representação gráfica que, quanto mais à direita no eixo das abscissas estiver o ponto de inflexão, mais difícil será considerado o item. Em outros termos, por se tratar da avaliação de desempenho, interpreta-se que os itens com valores maiores de b são mais difíceis de serem bem avaliados.

Ainda em relação ao parâmetro de dificuldade, Mello (2014) destaca que não se trata de um indicador de qualidade do teste, mas que demonstra em qual região na escala o item possui maior informação. A Tabela 8 apresenta os valores estimados para os parâmetros a e b de cada item do conjunto, destacando os menores e maiores valores encontrados.

Tabela 8 – Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens

Item	Assertiva	Parâmetro <i>a</i>	Parâmetro <i>b</i>
01	<i>O programa da disciplina foi apresentado pelo professor (objetivos, conteúdo a ser desenvolvido e bibliografia)</i>	2,257	-1,552
02	<i>O professor comparece às aulas e cumpre os horários de início e de término das mesmas</i>	1,881	-1,698
03	<i>O professor utiliza de forma adequada os recursos didáticos disponíveis, favorecendo a aprendizagem</i>	3,072	-1,053
04	<i>O professor estimula a utilização de materiais complementares (livros, sites, periódicos online, áudio, vídeos, entre outros)</i>	2,107	-1,094
05	<i>O professor demonstra domínio sobre o conteúdo apresentado, tratando-o com clareza e objetividade</i>	2,812	-1,249
06	<i>O professor estabelece relações entre os conteúdos da sua disciplina com os conteúdos das demais disciplinas, contribuindo com a formação profissional</i>	2,509	-1,075
07	<i>O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes</i>	3,723	-1,077
08	<i>O professor mostra-se disponível para tirar dúvidas sobre a disciplina e o conteúdo ministrado</i>	3,177	-1,333
09	<i>O professor ouve críticas, opiniões e sugestões referentes às suas aulas, mostrando-se aberto ao diálogo</i>	3,045	-0,945
10	<i>O professor apresenta uma postura de respeito mútuo, ou seja, preserva a imagem da Instituição, dos colegas, dos acadêmicos e não manifesta atitudes preconceituosas</i>	2,532	-1,499
11	<i>O professor elabora avaliações compatíveis com os conteúdos ministrados em aula</i>	2,537	-1,330
12	<i>Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula</i>	2,049	-0,850
13	<i>O professor cumpre o programa da disciplina apresentado</i>	2,680	-1,433

Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

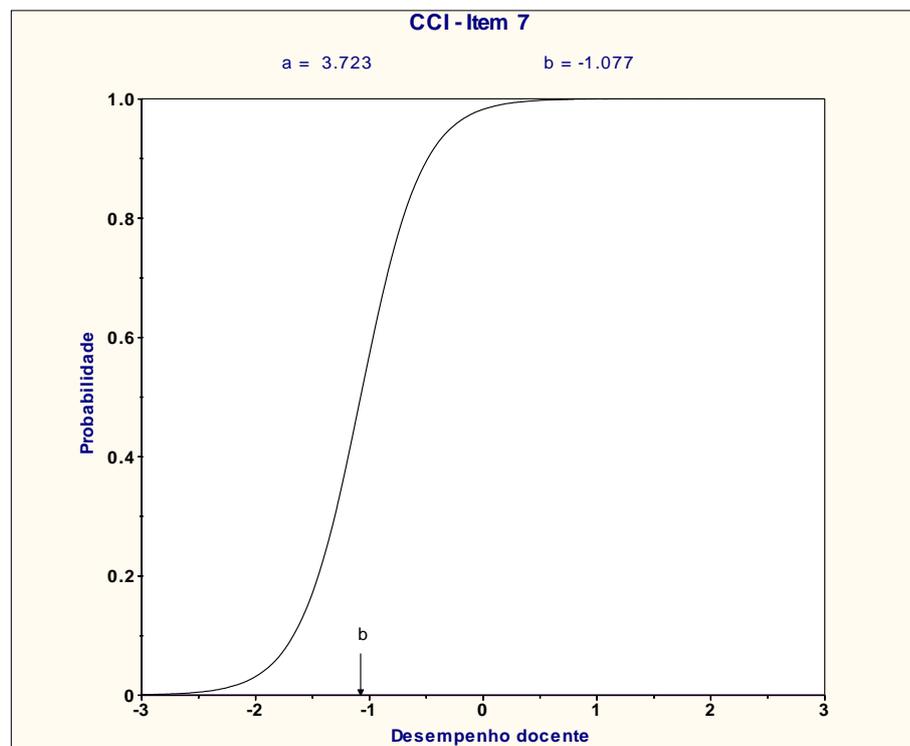
Em uma análise inicial, para fins de interpretação do parâmetro *a*, considera-se que itens com valores de $a < 1$ indicam que o item apresenta pouco poder de discriminação (VEY, 2011). Nas CCI, isto pode ser verificado a partir da inclinação obtida no ponto médio da curva em relação ao eixo *x*, em que os itens com maior poder de discriminação apresentam curvas mais íngremes, enquanto os demais apresentam curvas com ângulo de inclinação menor. Para o estudo em questão, o item 7 (“O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes”) foi o que resultou em melhor poder de discriminação ($a = 3,723$). Em contrapartida, o item 2 (“O professor comparece às aulas e cumpre os horários de início e de término das mesmas”) apresentou o valor de 1,881 para o referido parâmetro. Apesar disso, considerando que todos os itens apresentaram valor superior a 1 (um) para o parâmetro *a*, de maneira geral o conjunto de itens mostra-se bastante equilibrado, não havendo a necessidade de exclusão de nenhum item do construto.

Para interpretação do parâmetro b , não há referência a valores absolutos. Entretanto, considera-se que, quanto maior o valor de b na escala, maior será o grau de dificuldade para aprovação do desempenho docente na percepção dos discentes.

Com base nos dados da tabela, o item 12 (“Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula”) foi o que apresentou maior grau de dificuldade para ser bem avaliado. Sendo assim, este aspecto poderá ser trabalhado para que o desempenho seja melhorado em futuras avaliações. Por outro lado, o item 2 (“O professor comparece às aulas e cumpre os horários de início e de término das mesmas”) foi o que apresentou menor valor no parâmetro b , indicando que a percepção dos discentes para melhoria neste aspecto possui grau de dificuldade menor em relação aos demais itens do conjunto. Comparando os dois itens mencionados, verifica-se que é mais fácil o discente aprovar o desempenho docente no item 2 do que no item 12.

A Figura 7 apresenta a CCI do item 7 (“O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes”), identificado como o item com maior poder de discriminação ($a = 3,723$) em relação ao restante do conjunto.

Figura 7 – Curva Característica do Item 7

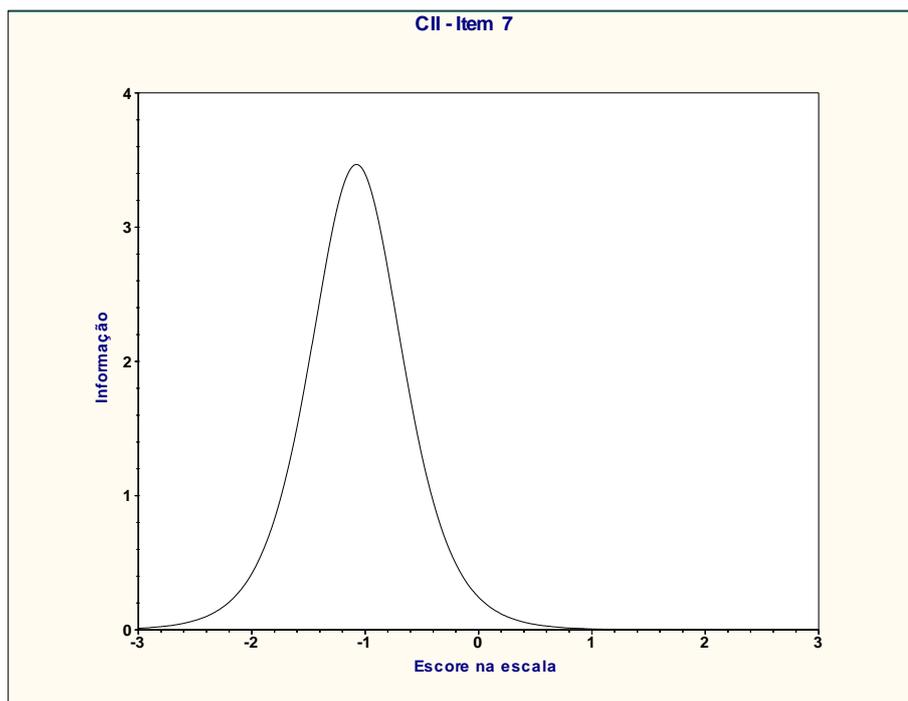


Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

Conforme mencionado anteriormente, observa-se um ângulo de inclinação bastante elevado, principalmente se comparada às CCI dos demais itens (vide Figura 5). Apesar disso, levando em consideração o valor do parâmetro b dos demais itens, este quesito ocupa apenas o quinto lugar no que diz respeito ao seu grau de dificuldade ($b = -1,077$). Infere-se que, em relação ao restante do conjunto de itens, os discentes da instituição encontraram alguma dificuldade em avaliar positivamente o desempenho docente.

Ainda em relação ao item 7, a análise gráfica das curvas de informação dos itens (CII) do teste indicou que este concentra a maior amplitude de informação em torno do nível de habilidade (θ), sendo fortemente influenciada pelo parâmetro de discriminação, conforme evidencia a Figura 8.

Figura 8 – Curva de Informação do Item 7



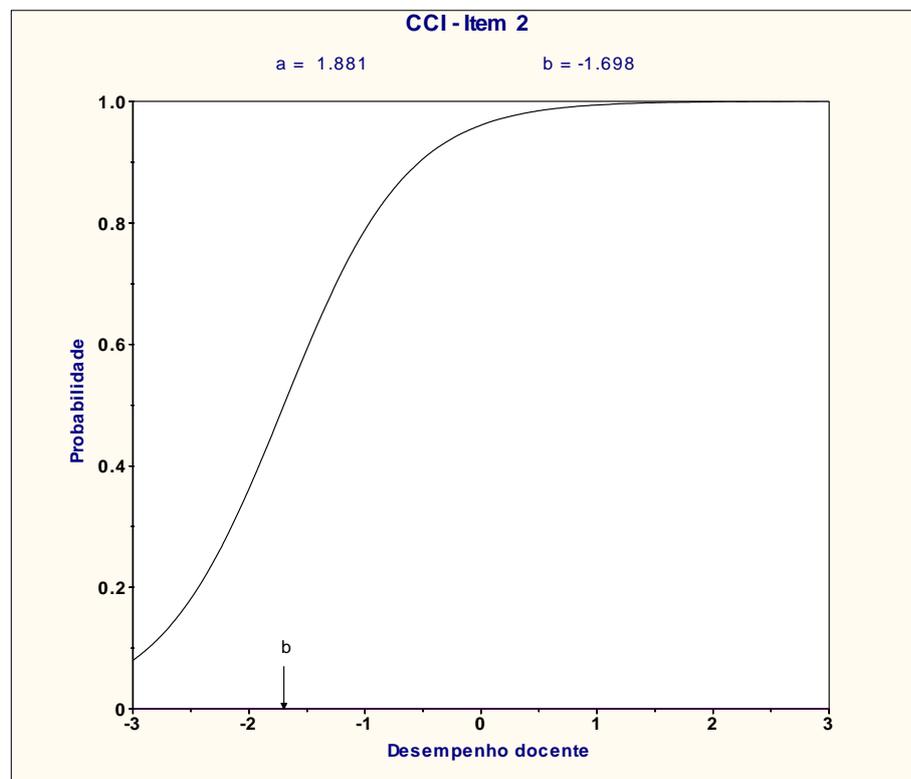
Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

A CII é baseada na função de informação do item (FII), e “reflete a qualidade dos itens individuais e do conjunto de itens como um todo” (VEY, 2011), empregados, neste caso, à avaliação de desempenho docente. Além disso, a FII pode auxiliar a identificar itens que apresentem baixo ou elevado grau de desempenho, podendo

indicar a má construção do item. Salienda Vey (2011) que pode haver, nestes casos, má formulação do enunciado de uma questão, ou que o mesmo se encontra fora do contexto do instrumento de medida como um todo.

Na Figura 9 está representada a curva característica do item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”), item que apresentou o menor valor em relação ao parâmetro a , 1,881. Apesar de o item ter apresentado o menor valor no parâmetro de discriminação em relação aos demais itens do conjunto, o fato de ter atingido um valor bastante superior a 1 (um) permite interpretá-lo como um item bem construído.

Figura 9 – Curva Característica do Item 2

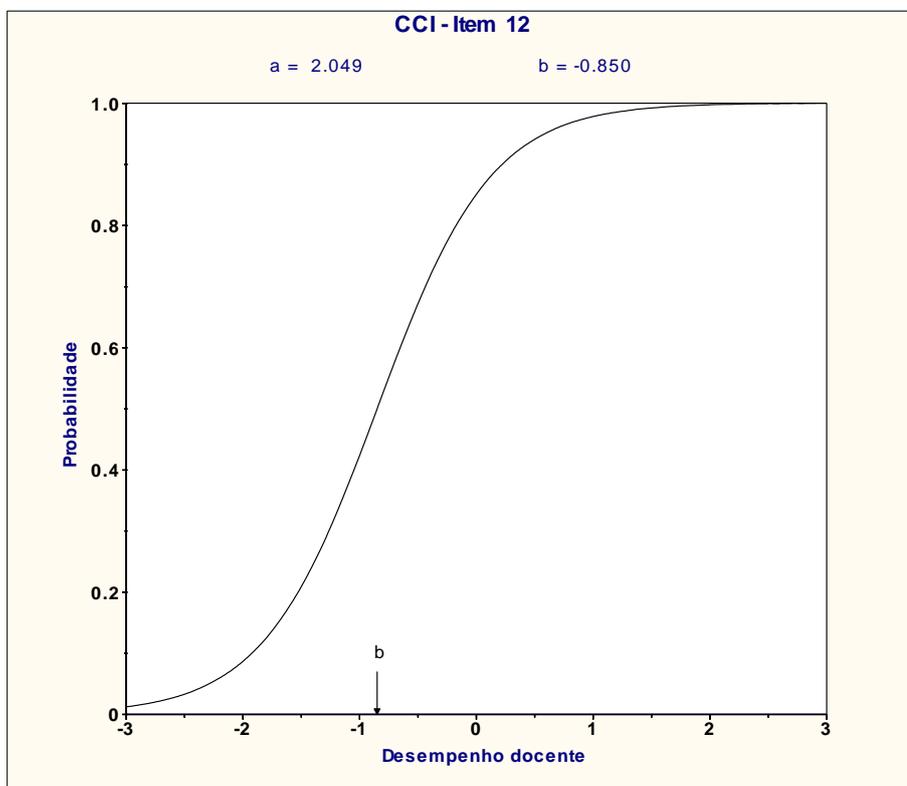


Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Coincidentemente, o item 2 foi o que apresentou também o menor valor no parâmetro b ($b = -1,698$), sendo, portanto, o item de menor dificuldade para que os respondentes concordassem com o desempenho do docente avaliado, ou seja, aprovassem seu desempenho docente.

O item 12 (“Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula”) foi o que apresentou maior valor do parâmetro b ($b = -0,850$), conforme evidencia a Figura 10. Portanto, neste quesito os acadêmicos encontraram maior dificuldade em concordar ou aprovar o desempenho docente. Por outro lado, este item é o penúltimo no conjunto de itens em poder de discriminação ($a = 2,049$). Apesar disso, o valor permite inferir que o item apresenta boa qualidade de construção ($a > 1$), conforme recomenda a literatura.

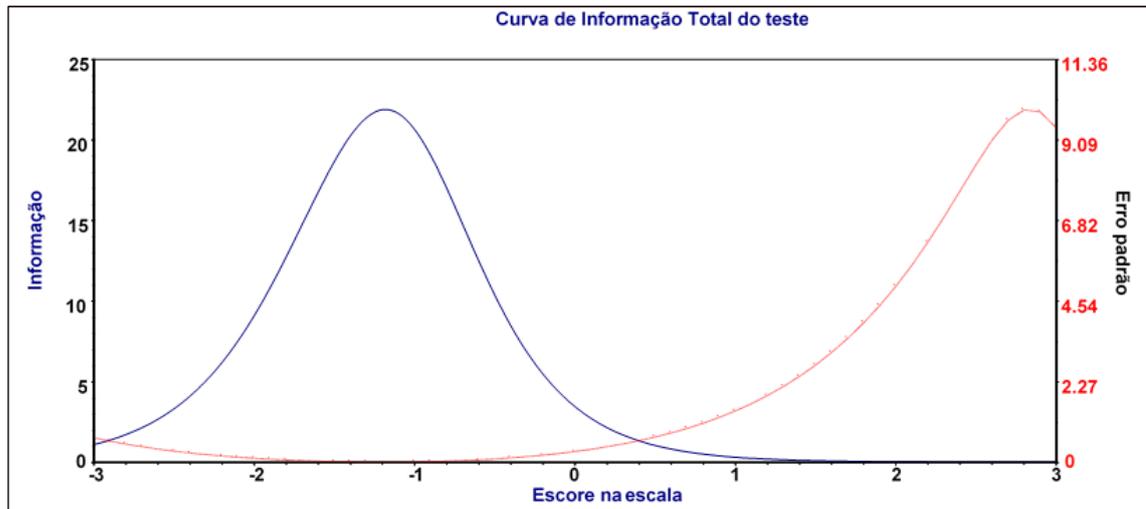
Figura 10 – Curva Característica do Item 12



Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

A Figura 11 representa a curva de informação total do teste. O gráfico apresenta o somatório da informação de todos os itens que compõem o teste, e é obtido a partir da função de informação do teste (FIT), demonstrando para qual região da escala o construto é mais indicado (VEY, 2011), se para avaliar docentes com desempenho baixo, médio ou elevado.

Figura 11 – Curva de Informação Total do teste



Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

No gráfico são apresentadas a função de informação do teste e o erro padrão de medida (EPM) para o conjunto de itens. A FIT é representada pela linha contígua, devendo ser analisada pela escala vertical situada à esquerda do gráfico. A linha pontilhada representa o EPM, e sua análise é feita a partir das informações constantes na escala vertical à direita.

Em relação à FIT, verificou-se que a maior quantidade de informação – acima de 10 – para os níveis de desempenho docente situa-se no intervalo de -1,9 a -0,5. Em relação ao EPM (linha pontilhada), interpreta-se que quanto maior a quantidade de informação, menor será o erro padrão e maior a precisão. No caso, o menor erro padrão também se situou no intervalo entre -1,9 e -0,5. O EPM aumenta para valores mais elevados na escala de desempenho docente, atingindo valor superior a 9 no extremo direito, para $\theta = 2,8$. Já no extremo esquerdo, o EPM atingiu valor superior a 1,5 para $\theta = -3$.

Considerando os resultados obtidos, verificou-se que o conjunto de itens do construto é mais adequado para medir itens entre os níveis -2,7 e 0,2 na escala (0, 1). Além disso, o formato da curva também é um indicador de validade do construto. Tendo em vista sua forma próxima a de uma distribuição normal, pode-se reforçar a validade do conjunto de itens.

4.3.2 Síntese dos resultados por Unidade Universitária

Esta etapa do estudo analisou os dados obtidos para os parâmetros de discriminação e de dificuldade dos itens em cada Centro de Ensino, para verificar se os resultados vão ao encontro daqueles obtidos para a Universidade como um todo. Para tanto, foram realizados os mesmos testes aplicados na seção anterior, adequando a sintaxe do *Bilog-MG* para os bancos de dados individualizados. Tendo em vista que em todas as Unidades Universitárias o parâmetro *a* teve resultados superiores a 1 (um), o que atesta a qualidade de construção dos itens, a abordagem enfatizou a análise do item com maior valor no referido parâmetro em cada Unidade.

4.3.2.1 Centro de Artes e Letras – CAL

O Centro de Artes e Letras da UFSM atualmente conta com sete cursos de graduação presenciais, totalizando 1.395 alunos, conforme dados do site “UFSM em números”. Alguns dos cursos estão disponíveis nas modalidades Licenciatura Plena e Bacharelado, como é o caso, por exemplo, dos cursos de Artes Visuais, ou então em especialidades diversas, como é o caso do curso de Música, com ênfase em determinados tipos de instrumentos. A Tabela 9 traz os cursos oferecidos pelo Centro de Artes e Letras, e o total de alunos por curso de graduação, conforme dados do ano de 2016.

Tabela 9 – Número de alunos por curso de graduação presencial – CAL

Curso	Nº de alunos
Artes Cênicas	82
Artes Visuais	218
Dança	41
Desenho Industrial	210
Letras	568
Música	218
Teatro	58
Total	1.395

Fonte: UFSM em Números.

Foram realizadas 1.551 avaliações por parte dos acadêmicos dos cursos do CAL, representando 6,50% de todas as avaliações consideradas no estudo. Conforme estudo realizado por Nunes e Primi (2005) para avaliar o impacto do tamanho da amostra na calibração dos itens, com base em um banco de dados com as respostas de 44.000 do Estado da Bahia, os autores concluíram que com amostras a partir de 200 participantes é possível estimar adequadamente os parâmetros dos itens e as habilidades dos respondentes. Portanto, o número de avaliações obtidas é suficientemente significativo para os resultados encontrados.

A Tabela 10 apresenta os valores dos parâmetros a e b obtidos para o Centro de Artes e Letras, destacando as células que contêm o maior e o menor valor em cada um dos parâmetros.

Tabela 10 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CAL

Item	Parâmetro a	Parâmetro b
01	2,389	-1,519
02	2,053	-1,754
03	4,084	-1,284
04	2,324	-1,439
05	3,164	-1,453
06	2,771	-1,262
07	4,109	-1,285
08	3,019	-1,578
09	3,191	-1,243
10	3,091	-1,512
11	3,829	-1,530
12	2,754	-1,209
13	2,832	-1,390

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

O item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”) apresentou o maior valor para o parâmetro de discriminação ($a = 4,109$), como já havia acontecido nos testes estatísticos envolvendo o conjunto agregado de todos os respondentes ($a = 3,723$), fazendo dele o item com maior poder de discriminação e indicando sua boa construção.

No que tange ao parâmetro b , o conjunto de itens oscilou de $b = -1,754$ no item 2 (“O professor comparece às aulas e cumpre os horários de início e de término das mesmas”) até $b = -1,209$ no item 12 (“Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula”), assim como havia acontecido com o resultado agregado.

4.3.2.2 Centro de Ciências Naturais e Exatas – CCNE

O Centro de Ciências Naturais e Exatas disponibiliza nove cursos de graduação presenciais à comunidade acadêmica, sendo alguns deles nas modalidades Licenciatura Plena e outros em Bacharelado. Em 2016, o CCNE teve 1.410 alunos matriculados nos cursos oferecidos pela Unidade Universitária (UFSM em Números). Na Tabela 11 estão detalhadas as informações de alunos por curso.

Tabela 11 – Número de alunos por curso de graduação presencial – CCNE

Curso	Nº de alunos
Ciências Biológicas	221
Estatística	113
Física	190
Geografia	263
Matemática	247
Meteorologia	66
Química	179
Química Industrial	52
Tecnologia em Proc. Químicos	79
Total	1.410

Fonte: UFSM em Números.

O CCNE foi a Unidade Universitária com maior percentual de avaliações respondidas em relação à instituição. As 4.314 avaliações representaram 18,09% do total captado. Em relação a isso, cabe lembrar que os departamentos do CCNE são responsáveis por atender parte do currículo básico de alguns cursos de Engenharia e da área de Informática, o que implica em alunos de cursos não vinculados ao CCNE avaliando docentes de seus departamentos.

Na Tabela 12 são apresentados os respectivos valores para os parâmetros a e b obtidos pelo Centro de Ciências Naturais e Exatas.

Tabela 12 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCNE

Item	Parâmetro a	Parâmetro b
01	2,171	-1,657
02	1,850	-1,795
03	3,170	-0,962
04	2,061	-0,973
05	2,752	-1,188
06	2,416	-0,935
07	3,774	-0,970
08	3,054	-1,304
09	3,354	-0,849
10	2,526	-1,503
11	2,225	-1,342
12	1,906	-0,941
13	2,549	-1,613

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Assim como ocorreu em relação ao Centro de Artes e Letras, o item 7 (“O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes”) apresentou valor do parâmetro a igual a 3,774, o que lhe conferiu o maior poder de discriminação em relação ao conjunto.

Em relação ao parâmetro de dificuldade, novamente o resultado apontou para o item 2 como sendo o de mais fácil aprovação ($b = -1,795$), assim como ocorreu no teste com os resultados agregados de todas as Unidades Universitárias. Disso inferiu-se que os acadêmicos têm pouca dificuldade em concordar com a afirmação “O professor comparece às aulas e cumpre os horários de início e de término das mesmas”. O item 9 (“O professor ouve críticas, opiniões e sugestões referentes às suas aulas, mostrando-se aberto ao diálogo”) foi considerado o mais difícil ($b = -0,849$) do conjunto, ou seja, o que apresentou maior dificuldade para concordância por parte do discente avaliador, em relação ao docente avaliado. Aqui observa-se uma diferença em relação ao resultado agregado obtido pelas Unidades Universitárias.

Enquanto o resultado agregado trouxe o item 12 como o mais difícil (“Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula”), na avaliação docente do CCNE este item ficou apenas em terceiro lugar ($b = -0,941$). Isto, no entanto, não invalida os resultados obtidos, apenas revela uma diferença de percepção entre os discentes que avaliaram docentes do Centro de Ciências Naturais e Exatas.

4.3.2.3 Centro de Ciências Rurais – CCR

O Centro de Ciências Rurais oferece atualmente cinco cursos de graduação, que atendem pouco mais de 1.900 alunos, conforme evidencia a Tabela 13. Com exceção do Curso Superior de Tecnologia em Alimentos, considerado tecnólogo, todos os demais são oferecidos na modalidade bacharelado.

Tabela 13 – Número de alunos por curso de graduação presencial – CCR

Curso	Nº de alunos
Agronomia	580
Engenharia Florestal	299
Medicina Veterinária	465
Tecnologia em Alimentos	261
Zootecnia	314
Total	1.919

Fonte: UFSM em Números.

O CCR ocupou a quarta posição em número de avaliações realizadas, com um total de 3.309, representando 13,87% de toda a amostra.

Na Tabela 14 são apresentados os resultados obtidos para os parâmetros a e b para cada um dos itens.

Tabela 14 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCR
(continua)

Item	Parâmetro a	Parâmetro b
01	2,548	-1,487
02	2,017	-1,589

Tabela 14 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCR

(conclusão)

Item	Parâmetro a	Parâmetro b
03	2,793	-1,052
04	2,199	-0,978
05	2,943	-1,215
06	2,857	-1,072
07	4,192	-0,982
08	3,305	-1,181
09	3,377	-0,840
10	3,054	-1,381
11	2,514	-1,188
12	2,182	-0,658
13	3,227	-1,372

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Os resultados se aproximaram daqueles obtidos no conjunto agregado de todas as Unidades Universitárias, tendo em vista que o item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”) apresentou o menor poder de discriminação e menor dificuldade ($a = 2,017$ e $b = -1,589$, respectivamente), assim como no resultado agregado (vide Tabela 8). Da mesma forma, o item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”) apresentou o melhor poder de discriminação para os dois conjuntos de dados ($a = 4,192$ para o CCR e $a = 3,723$ para os dados agregados). Também, repetese o resultado para o item considerado o mais difícil, no caso o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”), com valores de $b = -0,658$ para o CCR e $b = -0,850$ para o conjunto agregado com os dados de todas as Unidades Universitárias.

4.3.2.4 Centro de Ciências da Saúde – CCS

O Centro de Ciências da Saúde possui 2.260 alunos divididos em seus sete cursos, todos eles ofertados na modalidade bacharelado, conforme exposto na Tabela 15. A participação relativa do Centro de Ensino no total de avaliações foi de 12,45%,

o que totalizou 2.971 questionários respondidos. Este valor deixou o CCS em quinto lugar no número total de avaliações.

Tabela 15 – Número de alunos por curso de graduação presencial – CCS

Curso	Nº de alunos
Enfermagem	200
Farmácia	449
Fisioterapia	215
Fonoaudiologia	118
Medicina	659
Odontologia	319
Terapia Ocupacional	300
Total	2.260

Fonte: UFSM em Números.

A Tabela 16 contém os valores dos parâmetros de discriminação e dificuldade (a e b , respectivamente) obtidos a partir dos testes realizados no *software Bilog-MG*.

Tabela 16 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CCS

Item	Parâmetro a	Parâmetro b
01	2,302	-1,439
02	2,005	-1,555
03	3,078	-1,070
04	2,036	-1,153
05	2,403	-1,420
06	2,406	-1,177
07	3,611	-1,184
08	3,427	-1,311
09	3,135	-0,917
10	2,351	-1,461
11	2,494	-1,295
12	1,884	-0,549
13	2,654	-1,464

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Mais uma vez, o item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”) apresentou maior poder de discriminação, com $a = 3,611$, o que indicou sua boa qualidade de construção.

Com relação aos parâmetros de dificuldade, estes oscilaram entre $b = -1,555$ e $b = -0,549$, para os itens 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”) e 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) respectivamente, indo ao encontro do resultado obtido para o conjunto do teste com todas as Unidades Universitárias.

4.3.2.5 Centro de Ciências Sociais e Humanas – CCSH

O Centro de Ciências Sociais e Humanas é a Unidade Universitária com o maior número de alunos matriculados em cursos de graduação. Conforme evidencia a Tabela 17, em 2016 o CCSH atendeu mais de 3.500 alunos em seus treze cursos, a maioria na modalidade bacharelado. Apesar disso, o Centro possui cursos na modalidade Licenciatura Plena, além de oferecer cursos noturnos, como Administração e Ciências Contábeis. Em relação à participação na avaliação de desempenho docente, o CCSH ficou em terceiro lugar com 4.208 questionários respondidos, ficando atrás apenas do Centro de Ciências Naturais e Exatas e do Centro de Tecnologia, primeiro e segundo colocados, respectivamente.

Tabela 17 – Número de alunos por curso de graduação presencial – CCSH
(continua)

Curso	Nº de alunos
Administração	465
Arquivologia	131
Ciências Contábeis	392
Ciências Econômicas	317
Ciências Sociais	295
Comunicação Social	467
Direito	423
Filosofia	234
História	191
Psicologia	131

Tabela 17 – Número de alunos por curso de graduação presencial – CCSH
(conclusão)

Curso	Nº de alunos
Relações Internacionais	234
Serviço Social	230
Sociologia	6
Total	3.516

Fonte: UFSM em Números.

A Tabela 18 contém os resultados obtidos para os parâmetros de discriminação e dificuldade do conjunto de itens avaliados no CCSH.

Tabela 18 – Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens – CCSH

Item	Parâmetro a	Parâmetro b
01	2,417	-1,480
02	1,728	-1,718
03	3,021	-1,050
04	2,290	-1,121
05	2,918	-1,244
06	2,770	-1,043
07	3,694	-1,115
08	3,102	-1,436
09	2,677	-1,047
10	2,364	-1,512
11	2,850	-1,323
12	2,154	-0,978
13	3,018	-1,254

Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

No que diz respeito ao parâmetro de discriminação, o comportamento dos itens se mostrou semelhante ao resultado agregado. Em ambos os testes, o item 2 (“O professor comparece às aulas e cumpre os horários de início e de término das mesmas”) apresentou o menor poder de discriminação e o item 7 (“O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes”)

obteve o maior valor neste parâmetro. No CCSH, os valores de a oscilaram de $a = 1,728$ até $a = 3,694$, indicando boa qualidade de construção dos itens.

A análise do parâmetro de dificuldade também se aproximou do resultado obtido com os dados agregados, em que o item 2 foi considerado o mais fácil, e o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) o mais difícil ($b = -1,718$ e $b = -0,978$, respectivamente).

4.3.2.6 Centro de Educação – CE

O Centro de Educação conta com três cursos de graduação presencial, sendo que os cursos de Educação Especial e de Pedagogia são oferecidos nos turnos diurno e noturno. Já o Programa Especial de Graduação de Formação de Professores para a Educação Profissional funciona em horários diferenciados, com aulas concentradas às sextas-feiras e aos sábados. Este último trata-se de uma ação articulada da qual fazem parte também o Centro de Ciências Rurais e o Colégio Técnico Industrial, e tem por objetivo formar professores em nível superior para a docência na modalidade de educação profissional.

Nos três cursos citados, o Centro de Educação atendeu a um total de 1.210 alunos no ano de 2016, conforme explicita a Tabela 19. Em termos de participação na avaliação de desempenho, foram respondidos 1.739 questionários referentes aos docentes lotados no CE.

Tabela 19 – Número de alunos por curso de graduação presencial – CE

Curso	Nº de alunos
Educação Especial	380
Pedagogia	515
PEG de Formação de Professores para a Educação Profissional	315
Total	1.210

Fonte: UFSM em Números.

Na Tabela 20 são expostos os resultados encontrados para os parâmetros de discriminação (a) e dificuldade (b) dos itens, obtidos a partir dos testes rodados no

software Bilog-MG, utilizando-se a base contendo os dados dos respondentes no Centro de Educação.

Tabela 20 – Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens – CE

Item	Parâmetro <i>a</i>	Parâmetro <i>b</i>
01	2,267	-1,701
02	2,222	-2,067
03	3,210	-1,512
04	2,424	-1,825
05	2,991	-1,747
06	2,856	-1,375
07	3,713	-1,627
08	4,768	-1,585
09	3,457	-1,372
10	2,697	-1,822
11	3,282	-1,672
12	2,366	-1,202
13	3,089	-1,534

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Em relação ao parâmetro *a*, o item 8 (“*O professor mostra-se disponível para tirar dúvidas sobre a disciplina e o conteúdo ministrado*”) obteve o melhor poder de discriminação, com $a = 4,768$. Cabe mencionar que o Centro de Educação foi a única Unidade Universitária em que o item 8 apresentou o maior valor de *a*. Nos demais Centros de Ensino, o maior valor neste parâmetro foi obtido pelo item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”), com sete ocorrências de dez possíveis, assim como para o conjunto agregado das Unidades.

A análise do parâmetro de dificuldade revelou que o item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”) foi considerado o de mais fácil aprovação, com $b = -2,067$, e o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) foi considerado o mais difícil, com $b = -1,202$.

Na análise deste parâmetro, portanto, o resultado foi semelhante ao encontrado nos testes realizados com o conjunto total de respondentes.

4.3.2.7 Centro de Educação Física e Desportos – CEFD

O Centro de Educação Física e Desportos possui dois cursos de graduação presencial, tendo atendido em 2016 mais de 600 alunos. Dos cursos, o de Educação Física é ofertado nas modalidades Licenciatura Plena e Bacharelado, enquanto o de Dança é oferecido apenas na modalidade Licenciatura. Na Tabela 21 está indicado o número de alunos por curso.

Tabela 21 – Número de alunos por curso de graduação presencial – CEFD

Curso	Nº de alunos
Dança	54
Educação Física	573
Total	627

Fonte: UFSM em Números.

Tendo em vista estes números, a avaliação docente do CEFD contou com um total de 506 questionários respondidos, o que representou uma participação de 2,12% em relação ao total, deixando o Centro de Ensino em penúltimo lugar na participação percentual.

A análise dos parâmetros *a* e *b* obtidos pelo CEFD mostrou resultados diversos daqueles encontrados para o conjunto agregado, e mesmo dos resultados individuais das Unidades Universitárias, conforme pode ser observado na Tabela 22.

Tabela 22 – Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens – CEFD
(continua)

Item	Parâmetro <i>a</i>	Parâmetro <i>b</i>
01	3,029	-1,175
02	2,769	-1,295
03	3,611	-0,981
04	2,148	-1,196
05	2,985	-1,233

Tabela 22 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CEFD
(conclusão)

Item	Parâmetro a	Parâmetro b
06	2,430	-1,048
07	2,871	-1,204
08	2,745	-1,421
09	2,800	-0,988
10	1,556	-1,782
11	2,245	-1,460
12	2,232	-0,814
13	3,957	-1,008

Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

No que diz respeito ao parâmetro de discriminação, foi a primeira e única vez em que o item 10 (“*O professor apresenta uma postura de respeito mútuo, ou seja, preserva a imagem da Instituição, dos colegas, dos acadêmicos e não manifesta atitudes preconceituosas*”) apareceu no extremo inferior do comparativo com os demais itens, atingindo $a = 1,556$. Apesar do resultado, ressalta-se que ainda assim o item possui bom poder de discriminação. Na mesma linha, o item 13 (“*O professor cumpre o programa da disciplina apresentado*”) apresentou o melhor poder de discriminação ($a = 3,957$), diferentemente de todas as outras Unidades Universitárias.

A análise do parâmetro b também apresentou uma peculiaridade, se comparado aos valores encontrados para os demais Centros de Ensino. O item 10 foi considerado o mais fácil pela primeira vez, com $b = -1,782$. Resultado semelhante só foi encontrado na avaliação do Colégio Politécnico, cuja análise é abordada mais adiante. Considerado de mais difícil aprovação de desempenho por parte dos discentes, o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) teve parâmetro de dificuldade estimado em $b = -0,814$, seguindo a linha do resultado geral agregado.

4.3.2.8 Centro de Tecnologia – CT

O Centro de Tecnologia é uma das Unidades Universitárias com a maior variedade de cursos oferecidos, assim como o Centro de Ciências Sociais e Humanas.

No total, são quatorze cursos de graduação presencial, todos oferecidos na modalidade Bacharelado, que atenderam em 2016 a mais de 3.000 alunos, conforme os dados da Tabela 23.

Tabela 23 – Número de alunos por curso de graduação presencial – CT

Curso	Nº de alunos
Arquitetura e Urbanismo	164
Ciência da Computação	155
Engenharia Acústica	190
Engenharia Aeroespacial	72
Engenharia Civil	464
Engenharia de Computação	218
Engenharia de Controle e Automação	203
Engenharia de Produção	225
Engenharia de Telecomunicações	76
Engenharia Elétrica	322
Engenharia Mecânica	272
Engenharia Química	366
Engenharia Sanitária e Ambiental	200
Sistemas de Informação	139
Total	3.066

Fonte: UFSM em Números.

A participação do CT na avaliação de desempenho docente representou 17,80% do total de questionários respondidos, perfazendo 4.208 avaliações. Este resultado deixou a Unidade Universitária em segundo lugar no que tange à coleta de dados, ficando atrás apenas do Centro de Ciências Naturais e Exatas.

A Tabela 24 apresenta os resultados estimados para os parâmetros de discriminação e de dificuldade do teste aplicado ao Centro de Tecnologia.

Tabela 24 – Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens – CT

(continua)

Item	Parâmetro <i>a</i>	Parâmetro <i>b</i>
01	1,920	-1,588
02	1,537	-1,690

Tabela 24 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CT
(conclusão)

Item	Parâmetro a	Parâmetro b
03	2,809	-0,848
04	1,674	-0,866
05	2,512	-1,018
06	2,171	-0,952
07	3,131	-0,847
08	2,423	-1,210
09	2,658	-0,729
10	2,133	-1,455
11	2,121	-1,219
12	1,749	-0,632
13	2,132	-1,457

Fonte: elaborado pelo autor com resultados do software *Bilog-MG*.

A análise do parâmetro a revelou que o maior poder de discriminação foi estimado para o item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”), que obteve $a = 3,131$, seguindo a estimaco resultante do conjunto agregado de itens com todas as Unidades Universitrias.

A anlise do parmetro de dificuldade dos itens tambm mostrou resultados similares aos obtidos pelo conjunto que considera a totalidade de respondentes da Instituio. O valor deste parmetro oscilou de $b = -1,690$ para o item 2 (“*O professor comparece s aulas e cumpre os horrios de incio e de trmino das mesmas*”) at $b = -0,632$ para o item 12 (“*Aps as avaliaos, o professor apresenta os itens avaliados, esclarecendo a nota atribuda e discutindo as questos em sala de aula*”). Estes resultados para o CT se aproximam bastante do resultado agregado de todas os Centros de Ensino.

4.3.2.9 Colgio Tcnico Industrial – CTISM

O Colgio Tcnico Industrial de Santa Maria iniciou suas atividades no ano de 1967, oferecendo inicialmente os Cursos Tcnicos de Eletrotcnica e de Mecnica, vinculados ao Ensino Mdio. Atualmente, o CTISM conta com cursos tcnicos presenciais e de educao  distncia, alm de trs cursos de graduao presencial

e um programa de pós-graduação. Em 2016, no nível de graduação presencial, o CTISM contou com 307 alunos distribuídos entre os cursos de Tecnologia em Fabricação Mecânica e Tecnologia em Redes de Computadores, como explicita a Tabela 25. O Curso Superior de Tecnologia em Eletrônica Industrial terá início no primeiro semestre de 2017, não tendo, portanto, participação na avaliação de desempenho docente quando da realização deste estudo.

Tabela 25 – Número de alunos por curso de graduação presencial – CTISM

Curso	Nº de alunos
Tecnologia em Fabricação Mecânica	109
Tecnologia em Redes de Computadores	198
Total	307

Fonte: UFSM em Números.

Em relação à sua participação na pesquisa, o fato de o CTISM ter apresentado o menor número de avaliações (339 no total, representando 1,42% de todas as avaliações), não invalida os resultados aferidos (NUNES e PRIMI, 2005).

A Tabela 26 traz os valores estimados para os parâmetros de discriminação e dificuldade, respectivamente, obtidos a partir dos dados coletados pela avaliação realizada no CTISM.

Tabela 26 – Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens – CTISM

(continua)

Item	Parâmetro <i>a</i>	Parâmetro <i>b</i>
01	2,238	-2,099
02	1,300	-3,026
03	3,853	-1,506
04	3,030	-1,152
05	3,674	-1,376
06	3,052	-1,208
07	2,528	-1,639
08	2,900	-1,422
09	2,961	-1,388
10	1,803	-1,839

Tabela 26 – Parâmetros a (discriminação) e b (dificuldade) dos itens – CTISM

(conclusão)

Item	Parâmetro a	Parâmetro b
11	3,088	-1,556
12	3,537	-1,029
13	2,957	-1,683

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Em comparação aos resultados obtidos pelas demais Unidades Universitárias, destaca-se o fato de pela primeira vez o item 3 (“*O professor utiliza de forma adequada os recursos didáticos disponíveis, favorecendo a aprendizagem*”) ter aparecido com melhor poder de discriminação ($a = 3,853$), comparado aos demais itens do conjunto.

A análise do parâmetro b revelou o item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”) como o de mais fácil aprovação de desempenho ($b = -3,026$), e o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) como o de maior dificuldade, com $b = -1,029$. A análise conjunta dos parâmetros revela boa qualidade de construção dos itens, confirmando os resultados obtidos pelo conjunto agregado dos itens, mesmo no caso de uma amostra menor do que nas demais Unidades Universitárias.

4.3.2.10 Colégio Politécnico

O Colégio Politécnico da Universidade Federal de Santa Maria percorreu um longo caminho até adquirir suas atuais dimensões. O embrião que lhe deu origem foi a criação, em 1961, da Escola Agrotécnica de Santa Maria através da Lei 3.864-A de 24 de janeiro. A Escola ficou subordinada à Superintendência do Ensino Agrícola e Veterinária (BRASIL, 1961). Em 1968 a Escola foi transferida e vinculada à UFSM, oferecendo inicialmente o Curso Colegial Agrícola, equivalente ao Ensino Médio Técnico. A partir de 1996, o então Colégio Agrícola de Santa Maria (CASM) passa a oferecer cursos técnicos na modalidade Pós-Médio, como o Curso Técnico de Habilitação em Agropecuária, bem como o Técnico em Processamento de Dados e o Técnico em Informática. Apenas no ano de 2006, após a diversificação dos cursos oferecidos, o CASM se torna o Colégio Politécnico da UFSM. A oferta de cursos

superiores teve início em 2009, inicialmente com o Curso Superior de Tecnologia em Geoprocessamento e o Curso Superior de Tecnologia em Gestão de Cooperativas, e em 2010, com o também Curso Superior de Tecnologia em Sistemas para Internet.

Atualmente, o Politécnico oferece, além dos cursos superiores anteriormente citados, também o Curso de Graduação em Gestão Ambiental, incorporado da Unidade Descentralizada de Ensino Superior de Silveira Martins (UDESSM), que teve seus cursos distribuídos nas Unidades Universitárias do campus de Santa Maria. Na Tabela 27 encontra-se o número de alunos matriculados em cursos de graduação atendidos pelo Politécnico no ano de 2016.

Tabela 27 – Número de alunos por curso de graduação presencial – Politécnico

Curso	Nº de alunos
Tecnologia de Geoprocessamento	109
Tecnologia em Gestão de Cooperativas	159
Tecnologia em Sistemas para Internet	135
Total	403

Fonte: UFSM em Números.

A participação da Unidade Universitária contou com 670 questionários respondidos, correspondendo a 2,81% do total de avaliações coletadas.

A Tabela 28 contém os resultados estimados para os parâmetros a e b obtidos pelo Colégio Politécnico.

Tabela 28 – Parâmetros a (discriminação) e b (dificuldade) dos itens – Politécnico
(continua)

Item	Parâmetro a	Parâmetro b
01	2,349	-2,018
02	1,831	-2,024
03	2,957	-1,430
04	2,123	-1,572
05	2,693	-1,581
06	2,054	-1,418
07	4,833	-1,332
08	2,876	-1,958

Tabela 28 – Parâmetros a (discriminação) e b (dificuldade) dos itens – Politécnico
(conclusão)

Item	Parâmetro a	Parâmetro b
09	3,150	-1,370
10	2,595	-2,242
11	2,458	-1,824
12	2,657	-1,566
13	2,730	-1,779

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

A análise do parâmetro de discriminação se mostrou similar aos resultados obtidos por outras Unidades Universitárias, uma vez que os itens 2 (“O professor comparece às aulas e cumpre os horários de início e de término das mesmas”) e 7 (“O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes”) obtiveram o menor e o maior valor no parâmetro, respectivamente. O item 2 obteve $a = 1,831$, e o item 7 atingiu $a = 4,833$, formando os dois extremos na escala do parâmetro de discriminação.

Considerando que o item com menor valor no parâmetro ficou acima de um e próximo a dois, pode-se afirmar que o item apresenta boa qualidade ao diferenciar indivíduos com proficiências (θ) semelhantes. O mesmo se observa para o item 7, que atingiu o maior valor de a comparado aos resultados obtidos por todos os outros Centros de Ensino.

A análise do parâmetro b identificou uma singularidade em relação ao resultado das demais Unidades Universitárias, com o item 7 apresentando o maior grau de dificuldade pela primeira vez ($b = -1,332$). Apesar disso, comparando este resultado aos valores estimados para b nos outros Centros de Ensino, verifica-se que os valores guardam certa proporção, pois oscilaram de $b = -1,639$ até $b = -0,847$. De maneira semelhante ao ocorrido com o Centro de Educação Física e Desportos, o item 10 (“O professor apresenta uma postura de respeito mútuo, ou seja, preserva a imagem da Instituição, dos colegas, dos acadêmicos e não manifesta atitudes preconceituosas”) foi considerado o de mais fácil aprovação, com valor do parâmetro de dificuldade estimado em $b = -2,242$.

4.3.3 Comparativo dos resultados obtidos por Unidade Universitária

A comparação dos resultados obtidos pelas Unidades Universitárias quando da estimação dos parâmetros de discriminação e dificuldade (respectivamente *a* e *b*), revelou que há diferenças no que diz respeito à percepção dos discentes em relação ao desempenho docente nos diferentes Centros de Ensino. Esse fato, no entanto, não invalida o estudo, tendo em vista que o número da amostra foi diferente para cada Unidade, e que cada item individualmente atingiu os requisitos necessários para afirmar sua qualidade de construção.

A Tabela 29 agrupa os valores estimados para o parâmetro de discriminação de cada Unidade Universitária, ao mesmo tempo em que compara estes resultados com aqueles obtidos pelo conjunto agregado de todos os Centros. Pode-se assim identificar quais os melhores itens com base em cada amostra, qual a amplitude do parâmetro *a* em relação ao conjunto de itens, além de aferir em quais Unidades Universitárias os itens apresentaram maior ou menor poder de discriminação.

Tabela 29 – Comparativo do parâmetro de discriminação por Unidade Universitária

Item	UFSM	Unidades Universitárias									
		CAL	CCNE	CCR	CCS	CCSH	CE	CEFD	CT	CTISM	Politéc.
1	2,257	2,389	2,171	2,548	2,302	2,417	2,267	3,029	1,920	2,238	2,349
2	1,881	2,053	1,850	2,017	2,005	1,728	2,222	2,769	1,537	1,300	1,831
3	3,072	4,084	3,170	2,793	3,078	3,021	3,210	3,611	2,809	3,853	2,957
4	2,107	2,324	2,061	2,199	2,036	2,290	2,424	2,148	1,674	3,030	2,123
5	2,812	3,164	2,752	2,943	2,403	2,918	2,991	2,985	2,512	3,674	2,693
6	2,509	2,771	2,416	2,857	2,406	2,770	2,856	2,430	2,171	3,052	2,054
7	3,723	4,109	3,774	4,192	3,611	3,694	3,713	2,871	3,131	2,528	4,833
8	3,177	3,019	3,054	3,305	3,427	3,102	4,768	2,745	2,423	2,900	2,876
9	3,045	3,191	3,354	3,377	3,135	2,677	3,457	2,800	2,658	2,961	3,150
10	2,532	3,091	2,526	3,054	2,351	2,364	2,697	1,556	2,133	1,803	2,595
11	2,537	3,829	2,225	2,514	2,494	2,850	3,282	2,245	2,121	3,088	2,458
12	2,049	2,754	1,906	2,182	1,884	2,154	2,366	2,232	1,749	3,537	2,657
13	2,680	2,832	2,549	3,227	2,654	3,018	3,089	3,957	2,132	2,957	2,730

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Na tabela, cada coluna está preenchida com o valor resultante de *a* de cada item dentro de seu Centro de Ensino, com cada linha representando um item. A interpretação pode ser feita horizontalmente, em que as células grifadas em cinza

escuro representam o maior valor obtido para o item n no comparativo entre os Centros de Ensino, e as células grifadas em cinza claro representam o menor valor resultante. Verticalmente, pode-se verificar quais itens apresentaram maior e menor valor dentro de uma Unidade Universitária.

Inicialmente, pode-se verificar que o menor valor geral para o parâmetro de discriminação ocorreu no Colégio Técnico Industrial, em que o item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”) atingiu um valor de $a = 1,300$. Apesar de se tratar do menor valor encontrado, o fato de $a > 1$ justifica a manutenção do item no conjunto. Na extremidade oposta, o item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”) obteve $a = 4,833$ no resultado do Colégio Politécnico, ou seja, nesta Unidade o item 7 apresentou melhor poder de discriminação em relação a todas as demais.

Analisando-se item a item os resultados estimados para o parâmetro em cada Unidade Universitária, e comparando com o resultado agregado para a UFSM, verificou-se que, na maioria dos casos, os parâmetros das Unidades ficaram próximos aos da universidade, com algumas exceções. O item 3 (“*O professor utiliza de forma adequada os recursos didáticos disponíveis, favorecendo a aprendizagem*”) teve o parâmetro de discriminação estimado em $a = 3,072$ para a UFSM, e em $a = 4,084$ para o Centro de Artes e Letras, uma diferença de 32,94%. Nos demais casos, essa diferença não superou 26% para o item 3.

O item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”) apresentou parâmetro $a = 3,723$ para o conjunto agregado, e $a = 4,833$ para o Colégio Politécnico, uma diferença aproximada de 29,8%. A segunda maior diferença e todos os demais resultados do parâmetro de discriminação para o item 7 foram inferiores a 13%.

Merece destaque também o item 8 (“*O professor mostra-se disponível para tirar dúvidas sobre a disciplina e o conteúdo ministrado*”), que obteve um valor estimado de $a = 3,177$ para a UFSM, e $a = 4,768$ para o Centro de Educação, representando uma diferença de 50,08%. Nos demais casos, as diferenças foram inferiores a 8%.

O resultado da estimação com os dados agregados para o item 11 (“*O professor elabora avaliações compatíveis com os conteúdos ministrados em aula*”) revelou um valor de $a = 2,537$ para a Universidade. No CAL, o valor estimado para o parâmetro no mesmo item foi de $a = 3,829$, representando uma diferença de quase 51% deste

em relação àquele. A segunda maior diferença foi apresentada pelo CE, ficando em torno de 29%.

A maior disparidade entre os valores do parâmetro a comparando os resultados estimados para a Universidade, e aqueles obtidos pelos Centros de Ensino individualmente, ficou por conta do item 12 (*“Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula”*). Para a UFSM, o valor estimado para o parâmetro foi de $a = 2,049$, ante $a = 3,537$ obtido pelo Colégio Técnico Industrial. Isto representa uma diferença de quase 73% em relação ao valor obtido para a Instituição.

Por fim, o item 13 (*“O professor cumpre o programa da disciplina apresentado”*) também apresentou diferença significativa no caso do Centro de Educação Física e Desportos. Neste, o parâmetro de discriminação estimado para o item 13 foi de $a = 3,957$, diante de $a = 2,680$ para o resultado da UFSM, representando uma diferença de 47,65% em relação ao resultado dos dados agregados.

Apesar destas considerações, cabe mais uma vez ressaltar que os itens que compuseram o instrumento de medida apresentaram boa qualidade de construção, uma vez que em todos os casos o valor do parâmetro de discriminação foi superior a 1 (um), tanto pela análise do conjunto agregado contendo todos os respondentes, quanto para os resultados estimados para as Unidades Universitárias individualmente.

Análise similar foi efetuada compilando-se os dados obtidos para o parâmetro de dificuldade, permitindo identificar em quais Centros de Ensino os itens foram considerados menos ou mais difíceis de ter o desempenho docente aprovado, na percepção dos discentes participantes da avaliação. A Tabela 31 agrupa os valores estimados para o parâmetro de dificuldade de cada Centro de Ensino, e permite a comparação dos resultados tanto horizontal quanto verticalmente.

Tabela 30 – Comparativo do parâmetro de dificuldade por Unidade Universitária
(continua)

Item	UFSM	Unidades Universitárias									
		CAL	CCNE	CCR	CCS	CCSH	CE	CEFD	CT	CTISM	Politéc.
1	-1,552	-1,519	-1,657	-1,487	-1,439	-1,480	-1,701	-1,175	-1,588	-2,099	-2,018
2	-1,698	-1,754	-1,795	-1,589	-1,555	-1,718	-2,067	-1,295	-1,690	-3,026	-2,024
3	-1,053	-1,284	-0,962	-1,052	-1,070	-1,050	-1,512	-0,981	-0,848	-1,506	-1,430
4	-1,094	-1,439	-0,973	-0,978	-1,153	-1,121	-1,825	-1,196	-0,866	-1,152	-1,572
5	-1,249	-1,453	-1,188	-1,215	-1,420	-1,244	-1,747	-1,233	-1,018	-1,376	-1,581
6	-1,075	-1,262	-0,935	-1,072	-1,177	-1,043	-1,375	-1,048	-0,952	-1,208	-1,418

Tabela 30 – Comparativo do parâmetro de dificuldade por Unidade Universitária

(conclusão)

Item	UFSM	Unidades Universitárias									
		CAL	CCNE	CCR	CCS	CCSH	CE	CEFD	CT	CTISM	Politéc.
7	-1,077	-1,285	-0,970	-0,982	-1,184	-1,115	-1,627	-1,204	-0,847	-1,639	-1,332
8	-1,333	-1,578	-1,304	-1,181	-1,311	-1,436	-1,585	-1,421	-1,210	-1,422	-1,958
9	-0,945	-1,243	-0,849	-0,840	-0,917	-1,047	-1,372	-0,988	-0,729	-1,388	-1,370
10	-1,499	-1,512	-1,503	-1,381	-1,461	-1,512	-1,822	-1,782	-1,455	-1,839	-2,242
11	-1,330	-1,530	-1,342	-1,188	-1,295	-1,323	-1,672	-1,460	-1,219	-1,556	-1,824
12	-0,850	-1,209	-0,941	-0,658	-0,549	-0,978	-1,202	-0,814	-0,632	-1,029	-1,566
13	-1,433	-1,390	-1,613	-1,372	-1,464	-1,254	-1,534	-1,008	-1,457	-1,683	-1,779

Fonte: elaborado pelo autor com resultados do *software Bilog-MG*.

Verificou-se que o item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”), no Colégio Técnico Industrial obteve $b = -3,026$, sendo, portanto, considerado o item de mais fácil aprovação de desempenho no comparativo com todos os demais resultados, segundo a percepção dos discentes dos CTISM. O Centro de Ensino em que este item foi considerado mais difícil foi o CEFD, que obteve valor estimado para $b = -1,295$.

O item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) foi considerado o mais difícil de todo o conjunto de itens ($b = -0,549$). Esse valor foi obtido pelo Centro de Ciências da Saúde.

A análise item a item dos resultados estimados para o parâmetro de dificuldade também revelou algumas oscilações ao comparar-se os resultados das Unidades em relação ao resultado da UFSM. Para fins de análise, serão destacados apenas os casos em que a variação foi superior a 50%.

O primeiro caso a apresentar esta característica foi encontrado no Colégio Técnico Industrial, para os resultados do item 2 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”). A estimação do parâmetro para a UFSM resultou em $b = -1,698$, e para o CTISM $b = -3,026$, uma diferença de mais de 78% deste em relação àquele.

O item 4 (“*O professor estimula a utilização de materiais complementares (livros, sites, periódicos online, áudio, vídeos, entre outros)*”) apresentou uma diferença de 66,82% comparando-se o resultado obtido pelo Centro de Educação, em relação ao resultado obtido pela Universidade ($b = -1,825$ e $b = -1,094$).

Para o item 7 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”), a estimação do parâmetro de dificuldade para a UFSM resultou em $b = -1,077$. A maior disparidade ao comparar com os valores das Unidades Universitárias foi encontrada no CTISM ($b = -1,639$), representando uma diferença de 52,18% no resultado deste Centro de Ensino, em relação ao conjunto agregado dos itens.

A última variação a ser destacada envolve o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”), que apresentou o maior valor de b para o conjunto agregado de todas as Unidades, sendo, portanto, considerado o de maior dificuldade para aprovação do desempenho docente. Nesta situação, o valor estimado para este item no cômputo geral dos dados da avaliação foi de $b = -0,850$, ante $b = -1,556$ estimado para o Colégio Politécnico, representando uma diferença de mais de 84% em relação ao resultado agregado.

5 CONCLUSÕES E RECOMENDAÇÕES

Este trabalho teve como objetivo analisar o comportamento do conjunto de itens utilizado no instrumento de medida aplicado à avaliação docente pelos discentes da Universidade Federal de Santa Maria, utilizando a Teoria da Resposta ao Item.

Assim, foram estabelecidos seis objetivos específicos: 1) verificar a consistência interna, a confiabilidade e a dimensionalidade do instrumento de medida aplicado aos discentes; 2) verificar a unidimensionalidade do instrumento através da aplicação da análise fatorial de informação plena; 3) analisar a correlação bisserial dos itens integrantes do teste; 4) verificar os parâmetros de discriminação e de dificuldade dos itens; 5) analisar o comportamento dos parâmetros a e b nos Centros de Ensino, comparando-os com o resultado geral da UFSM; e 6) propor melhorias para o instrumento de avaliação, caso fossem verificadas inconsistências.

Para que os objetivos fossem alcançados, a pesquisa iniciou-se com o levantamento bibliográfico sobre os temas centrais deste trabalho: a avaliação da educação superior no Brasil; os conceitos e pressupostos fundamentais das avaliações de desempenho; e os conceitos e aplicações da Teoria da Resposta ao Item. Elaborado o referencial teórico, e já de posse da base de dados da avaliação docente aplicado em caráter experimental durante o primeiro semestre do ano de 2016, iniciou-se o processo de adequação do mesmo para exportação dos dados para os *softwares* estatísticos utilizados na pesquisa. Em um primeiro momento, efetuou-se a análise com base na Teoria Clássica dos Testes com o auxílio dos *softwares MS-Excel* e *SPSS*. Em seguida, procedeu-se à análise com base na TRI com os resultados gerados pelos programas *TESTFACT* e *Bilog-MG*.

A análise baseada na TCT buscou avaliar a qualidade do instrumento de medida verificando sua fidedignidade e dimensionalidade. Com esta finalidade, o primeiro teste executado no banco de dados da avaliação docente buscou determinar o alfa de Cronbach do conjunto de itens. O resultado do teste alfa de Cronbach para o instrumento de medida foi de 0,94, atestando que o conjunto de itens possui confiabilidade e consistência interna, atendendo, dessa forma, parte do primeiro objetivo de pesquisa.

Em sequência, o conjunto de itens foi carregado novamente no programa *SPSS* para o teste de dimensionalidade, para o qual foi utilizado a análise fatorial de componentes principais. O resultado revelou a existência de um fator determinante

responsável por 60,10% da variabilidade dos dados, com o segundo fator respondendo por apenas 6,65% da variância total. Estes resultados atestam que o instrumento de medida é unidimensional, ou seja, mede um único traço latente, neste caso, o desempenho docente. Assim, conclui-se o primeiro objetivo específico da pesquisa. Concluída a análise baseada na TCT, procedeu-se à realização dos testes específicos para a Teoria da Resposta ao Item.

Para a análise baseada na TRI, o banco de dados da avaliação docente foi dicotomizado. As respostas ao questionário encontravam-se em uma escala de concordância tipo Likert de 5 pontos. A dicotomização consistiu em converter as respostas 1, 2 e 3 em “não aprova o desempenho docente” ($x = 0$), e as respostas 4 e 5 em “aprova o desempenho docente” ($x = 1$). A seguir, a base de dados dicotomizada foi carregada no *software TESTFACT* para que fosse executada a análise fatorial de informação plena, também para verificar a dimensionalidade do instrumento. O resultado obtido foi semelhante ao da análise fatorial pelo método de componentes principais, mas neste caso o primeiro fator foi responsável por explicar 63,70% da variabilidade geral dos dados, e o segundo apenas 4,52%. Comprova-se assim a suposição inicial de que o conjunto de itens é unidimensional, alcançando-se o segundo objetivo específico.

A análise da correlação bisserial tem como objetivo verificar se determinado item binário guarda correlação significativa com o escore bruto (SOARES e MENDONÇA, 2003). No entendimento dos autores, trata-se de uma medida estatística da capacidade de discriminação do item. Nesse sentido, Soares (2005) expõe que é usual a aceitação de valores superiores a 0,3 para este parâmetro. Os resultados atingidos pelo conjunto de itens variaram de 0,7860 até 1,0750, tratando-se de mais um indício de que os itens que compõem o instrumento de coleta de dados foram bem construídos, ao apresentarem forte consistência interna e correlação com o escore bruto. Atingiu-se, assim, o terceiro objetivo específico.

A próxima fase consistiu na estimação dos parâmetros de discriminação e dificuldade dos itens (a e b , respectivamente), etapa também conhecida como calibração (VEY, 2011). Para tanto, foi utilizado o método da Máxima Verossimilhança Marginal (MVM), e a convergência dos dados também foi testada pelos algoritmos EM (*Expectation – Maximization*) e NR (Newton-Raphson), no intuito de garantir a boa estimação dos parâmetros (MELLO, 2014). A convergência do conjunto de itens para a UFSM foi atingida em 14 ciclos pelo algoritmo EM, e em 16 ciclos pelo algoritmo NR.

A análise do parâmetro de discriminação do conjunto de itens revelou valores de $a = 1,881$ até $a = 3,723$, indicando o bom poder de discriminação do conjunto de itens. A análise dos resultados destacou os dois itens que apresentaram o menor e o maior valor no parâmetro. Tem-se aqui mais um indício de qualidade do conjunto de itens, que reforça as constatações dos três primeiros objetivos específicos, assim como atende ao quarto.

Os resultados dos parâmetros de dificuldade do teste oscilaram entre $b = -1,698$ e $b = -0,850$. Apesar de não haver referência a valores absolutos, interpreta-se que, quanto maior o valor de b na escala, maior será o grau de dificuldade do item. Neste estudo, a leitura a ser feita é de que, quanto maior o valor de b na escala, maior a dificuldade para o discente aprovar o desempenho docente. A análise destacou os itens que apresentaram o menor e o maior valor no parâmetro de dificuldade.

A última etapa envolveu a estimação dos parâmetros dos itens em cada Centro de Ensino, e sua comparação com o resultado gerado para a UFSM como um todo. Para tanto, foi necessário novo tratamento ao banco de dados, que foi subdividido por Unidades Universitárias para carregamento no *Bilog-MG*. Com os parâmetros dos itens estimados para cada Centro, comparou-se os resultados obtidos para a e b nas Unidades em relação aos valores resultantes na UFSM. O comparativo permitiu verificar pequenas variações na calibração dos itens, e identificar os Centros de Ensino em que as variações em cada um dos parâmetros foram maiores. Apesar disso, as análises demonstraram que, mesmo nas Unidades com menor representatividade quanto ao número de respondentes, nenhum item deixou de satisfazer as condições para ser considerado de qualidade. Com base nisso, buscou-se atingir o quinto objetivo da pesquisa.

Os resultados encontrados neste trabalho a partir da aplicação da Teoria da Resposta ao Item, demonstram a versatilidade da técnica estatística ao proporcionar uma grande quantidade de informações e aplicações em diversas áreas de estudo. A partir destes resultados, pode-se afirmar que o problema de pesquisa foi respondido, ou seja, o instrumento de medida utilizado na avaliação docente pelos discentes apresentou confiabilidade e consistência interna, bem como atendeu aos pressupostos da unidimensionalidade e da independência local. Além disso, a etapa de calibração revelou que nenhum dos itens precisou ser desconsiderado, atestando assim a qualidade do construto.

Como contribuição do trabalho, destaca-se a aplicação da TRI na difusão de um tema que merece atenção e que não deve ser negligenciado, a avaliação de desempenho docente. Espera-se que este estudo proporcione discussões tanto a respeito da avaliação docente, quanto de outros serviços e segmentos da Universidade, dada a variedade de aplicações que a TRI permite. Por outro lado, espera-se que o trabalho ajude a consolidar a avaliação de desempenho, e que os resultados dela oriundos sejam utilizados para fins gerenciais no futuro, proporcionando *feedback* aos docentes para que gradualmente a qualidade das aulas ministradas aumente, na percepção discente. Almeja-se também que eventuais discussões decorrentes do trabalho fortaleçam a atuação das Comissões Setoriais de Avaliação junto aos Centros de Ensino.

Ficam como sugestões para novas pesquisas, a utilização dos modelos politômicos da TRI na análise dos dados da avaliação docente, assim como a aplicação da metodologia para avaliar os instrumentos de medida respondidos pelos estudantes dos demais níveis de ensino.

REFERÊNCIAS

ALEXANDRE, J. W. C. et al. Aplicação da teoria da resposta ao item na gestão da qualidade: proposta de um modelo probabilístico. In: XXI Encontro Nacional de Engenharia de Produção, 2001, Salvador. **Anais...** Disponível em: <http://www.abepro.org.br/biblioteca/ENEGEP2001_TR21_0036.pdf>. Acesso em: 12 nov. 2016.

ALMEIDA JUNIOR, V. O processo de implementação das Comissões Próprias de Avaliação (CPAs): ações desenvolvidas e perfil dos coordenadores. In: RISTOFF, D.; ALMEIDA JUNIOR, V. (Orgs.). **Avaliação participativa, perspectivas e desafios**. Brasília: INEP, 2005. p. 39-56.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item: conceitos e aplicações**. SINAPE, 2000.

ARAÚJO, E. A. C.; ANDRADE, D. F.; BORTOLOTTI, S. L. V. Teoria da resposta ao item. **Revista da Escola de Enfermagem da USP**, v.43, n. Esp, p. 1000-1008, dez. 2009.

AVES, M. L. F.; SOUZA, R. M. Q. A autoavaliação institucional no ensino superior: a perspectiva dos docentes e discentes. **Cadernos de Educação**, v. 12, n. 24, p. 232-234, jan.-jun. 2013.

AVRICHIR, I.; DEWES, F.. Construção e validação de um instrumento de avaliação do desempenho docente. In: Encontro da ANPAD, 30., 2006, Salvador. **Anais...** Salvador: 2006. Disponível em: <<http://www.anpad.org.br/enanpad/2006/dwn/enanpad2006-epqa-3002.pdf>>. Acesso em: 15 ago. 2016.

BALBIM JUNIOR, A. **Avaliação da satisfação de clientes de bancos utilizando a teoria da resposta ao item**. Florianópolis; 2010. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina.

BARBOSA, G. C.; FREIRE, F. S.; CRISÓSTOMO, V. L. Análise dos indicadores de gestão das IFES e o desempenho discente no ENADE. **Avaliação**, Campinas; Sorocaba, SP, v. 16, n. 2, p. 317-344, jul. 2011.

BARREYRO, G. B.; ROTHEN, J. C. Para uma história da avaliação da educação superior brasileira: análise dos documentos do PARU, CNRES, GERES e PAIUB. **Avaliação**, Campinas; Sorocaba, SP, v. 13, n. 1, p. 131-152, mar. 2008.

BOCLIN, R.. Avaliação de docentes do ensino superior: um estudo de caso. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 12, n. 45, p. 959-980, out.-dez. 2004.

BORTOLOTTI, S. L. V. et al. Avaliação do nível de satisfação de alunos de uma instituição de ensino superior: uma aplicação da teoria da resposta ao item. **Gestão & Produção**, São Carlos, v. 19, n. 2, p. 287-302, 2012.

BRASIL. Decreto n. 6.096 de 24 de abril de 2007. Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6096.htm>. Acesso em: 15 mai. 2016.

_____. Decreto n. 8.581 de 3 de dezembro de 2016. Altera o Decreto no 8.456, de 22 de maio de 2015, que dispõe sobre a programação orçamentária e financeira e estabelece o cronograma mensal de desembolso do Poder Executivo para o exercício de 2015. Disponível em: <http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2015/Decreto/D8581.htm>. Acesso em: 23 jun. 2016

_____. Lei n. 3.864-A de 24 de janeiro de 1961. Cria as Escolas Agrícolas de Bambuí e Cuiabá, nos Estados de Minas Gerais e Mato Grosso, e uma Escola de Engenharia em Uberlândia, Minas Gerais. Disponível em: <<http://www2.camara.leg.br/legin/fed/lei/1960-1969/lei-3864-a-24-janeiro-1961-353630-publicacaooriginal-1-pl.html>>. Acesso em: 02 fev. 2017.

_____. Lei n. 10.861 de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.861.htm>. Acesso em: 15 set. 2015.

_____. Portaria n. 2.051 de 9 de julho de 2004. Regulamenta os procedimentos de avaliação do Sistema Nacional de Avaliação da Educação Superior (SINAES), instituído na Lei n. 10.861, de 14 de abril de 2004. Disponível em: <http://portal.mec.gov.br/arquivos/pdf/PORTARIA_2051.pdf>. Acesso em: 16 set. 2015.

CASTANHEIRA, A. M.; CERONI, M. R. Reflexões sobre o processo de avaliar docente contribuindo com sua formação. **Avaliação**, Campinas; Sorocaba, SP, v.12, n. 4, p. 719-737, dez. 2007.

CASTRO, S. M. J. et al. Funcionamento diferencial do item no inventário de depressão de Beck. **Revista Brasileira de Epidemiologia**, São Paulo, v. 18, n. 1, p. 54-67, jan.-mar. 2015.

CERVO, A. L.; BERVIAN, P. A.; SILVA, R. **Metodologia científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007.

CHIAVENATO, I. **Recursos humanos: o capital humano das organizações**. 9. ed. Rio de Janeiro: Elsevier, 2009.

COUTO, G.; PRIMI, R. Teoria de Resposta ao Item (TRI): conceitos elementares dos modelos para itens dicotômicos. **Boletim de Psicologia**, São Paulo, v. 61, n. 134, p. 1-15, 2011.

DESSLER, G. **Administração de recursos humanos**. 2. ed. São Paulo: Prentice Hall, 2003.

EMBIRUÇU, M.; FONTES, C.; ALMEIDA, L. Um indicador para a avaliação do desempenho docente em Instituições de Ensino Superior. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 18, n. 69, p. 795-820, out.-dez. 2010.

FÁVERO, M. L. A. A universidade no Brasil: das origens à reforma universitária de 1968. **Educar**, Curitiba, n. 28, p. 17-36, 2006.

FIDELIS, J. F.; BANOV, M. R. **Gestão de recursos humanos: tradicional e estratégica**. 2. ed. São Paulo: Érica, 2007.

FRANCISCO, R. **Aplicação da teoria da resposta ao item (TRI) no exame nacional de cursos (E.N.C.) da UNICENTRO**. Curitiba, 2005. Dissertação (Mestrado em Estatística) – Universidade Federal do Paraná.

FRANÇA, A. C. L. **Práticas de recursos humanos – PRH: conceitos, ferramentas e procedimentos**. 1. ed. São Paulo: Atlas, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

_____. **Gestão de pessoas: enfoque nos papéis organizacionais**. 1. ed. São Paulo: Atlas, 2011.

HAIR JUNIOR, J. F. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005.

HOFFMANN, J. **Avaliação mediadora: uma prática em construção da pré-escola à universidade**. 20. ed. Porto Alegre: Mediação, 2003.

INEP. **Avaliação dos cursos de graduação**. Brasília, DF, 2015. Disponível em: <<http://portal.inep.gov.br/superior-condicoesdeensino>>. Acesso em: 09 nov. 2015.

_____. **Avaliação institucional**. Brasília, DF, 2015. Disponível em: <http://portal.inep.gov.br/superior-avaliacao_institucional>. Acesso em: 14 nov. 2015.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

_____. **Metodologia do trabalho científico**. 7. ed. São Paulo: Atlas, 2011.

LACOMBE, F. J. M. **Recursos humanos: princípios e tendências**. 2. ed. São Paulo: Saraiva, 2011.

MARRAS, J. P. **Administração de recursos humanos: do operacional ao estratégico**. 14. ed. São Paulo: Saraiva, 2011.

MELLO, L. A. **Validação de um instrumento de avaliação de desempenho de coordenadores de curso pela Teoria da Resposta ao Item**. Santa Maria; 2014. Dissertação (Mestrado em Gestão Pública) – Universidade Federal de Santa Maria.

MOREIRA JUNIOR, F. J. et al. Algoritmo de um teste adaptativo informatizado com base na teoria da resposta ao item para a estimação da usabilidade de sites de e-commerce. **Produção**, São Paulo, v. 23, n. 3, p. 525-536, jul.-set. 2013

MOTTA, F. C. P. O estruturalismo na Teoria das Organizações. **Revista de Administração de Empresas**, Rio de Janeiro, v.10, n. 4, p. 23-41, out.-dez. 1970.

NUNES, C. H. S. S.; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por Teoria da Resposta ao Item. **Avaliação Psicológica**, Itatiba, SP, v. 4, n. 2, p. 141-153, 2005.

NUNNALLY, J. C. **Psychometric theory**. 2. ed. New York: McGraw-Hill, 1978.

PASQUALI, L. Teoria da Resposta ao Item – IRT: uma introdução. In: PASQUALI, Luiz (org.). **Teoria e métodos de medida em ciências do comportamento**. Brasília: INEP, 1996. p. 173-195.

_____. **Psicometria: teoria dos testes na Psicologia e na Educação**. 5. ed. Petrópolis: Vozes, 2013.

PASQUALI, L.; PRIMI, R. Fundamentos da Teoria da Resposta ao Item – TRI. **Avaliação Psicológica**, Itatiba, SP, v. 2, n. 2, p. 99-110, 2003.

PEIXOTO, M. C. L. A avaliação institucional nas universidades federais e as comissões próprias de avaliação. **Avaliação**, Campinas; Sorocaba, SP, v. 14, n. 1, p. 9-28, mar. 2009.

RIBEIRO, J. L. L. S. SINAES: o que aprendemos acerca do modelo adotado para a avaliação do ensino superior no Brasil. **Avaliação**, Campinas; Sorocaba, SP, v. 20, n. 1, p. 143-161, mar. 2015.

RODRIGUES, S. S. Políticas de avaliação docente: tendências e estratégias. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, RJ, v. 20, n. 77, p. 749-768, out.-dez. 2012.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. Florianópolis: UFSC/PPGEP/LED, 2000.

SILVA, A. L.; GOMES, A. M. Avaliação institucional no contexto do SINAES: a CPA em questão. **Avaliação**, Campinas; Sorocaba, SP, v. 16, n. 3, p. 573-601, nov. 2011.

SILVA FILHO, G. M. **Nível de conhecimento dos profissionais de Contabilidade em relação ao sistema público de escrituração digital por meio da Teoria da Resposta ao Item**. João Pessoa; 2015. Dissertação (Mestrado em Ciências Contábeis) – UnB; UFPB; UFRN.

SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores socioeconômicos. **Pesquisa Operacional**, Rio de Janeiro, RJ, v.25, n. 1, p. 83-112, jan.-abr. 2005.

SOARES, T. M.; MENDONÇA, M. C. M. Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000. **Pesquisa Operacional**, Rio de Janeiro, RJ, v.23, n.3, p. 421-441, set.-dez. 2003.

SOUSA, A. C. G.; ANDRIOLA, W. B. Avaliação docente na educação superior: expectativas acadêmicas acerca da influência do feedback e das decisões institucionais. **Revista Eletrônica Acta Sapientia**, Fortaleza, CE, v.1, n. 1, 2014. Disponível em: <<http://www.poleduc.ufc.br/revista/index.php/actasap/article/view/2>>. Acesso em: 01 set. 2016.

TAVARES, H. R. **Teoria da Resposta ao Item para dados longitudinais**. Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2001.

TEZZA, R.; BORNIA, A. C. Teoria da Resposta ao Item: vantagens e oportunidades para a Engenharia de Produção. XXIX Encontro Nacional de Engenharia de Produção. **Anais...** Salvador: 2009.

UNIVERSIDADE FEDERAL DE SANTA MARIA. **Plano de Desenvolvimento Institucional 2011-2015**. Disponível em: <<http://site.ufsm.br/arquivos/uploaded/arquivos/be1eb2e0-4629-442e-b1af-79c251e3ac83.pdf>>. Acesso em: 16 nov. 2015.

_____. **Projeto de Avaliação Institucional**. Disponível em: <<http://coral.ufsm.br/cpa/images/documentos/projetoavaliacao.pdf>>. Acesso em: 16 nov. 2015.

VARGAS, V. C. C. et al. Avaliação dos intangíveis: uma aplicação em capital humano. **Gestão & Produção**, São Carlos, v. 15, n. 3, p. 619-634, set.-dez. 2008.

VARGAS, V. C. C. **Medida padronizada para avaliação de intangíveis organizacionais por meio da teoria da resposta ao item**. Florianópolis; 2007.

Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina.

VEY, I. H. **Avaliação de desempenho logístico no serviço ao cliente baseada na Teoria da Resposta ao Item**. Florianópolis; 2011. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina.

ZANDEVALLI, C. B. Avaliação da educação superior no Brasil: Os antecedentes históricos do SINAES. **Avaliação**, Campinas; Sorocaba, v.14, n. 2, p. 385-438, jul. 2009.

ANEXOS

ANEXO A – Questionário de avaliação docente pelo discente

Visualização Completa do Questionários

Informações do Questionário	
Programa Avaliação do Docente pelo Discente - 201601	Questionário Cursos de Graduação
Descrição do Programa Este programa tem como objetivo a aplicação de instrumento que visa identificar a percepção do estudante em relação à atuação dos professores nas atividades desenvolvidas no curso, com vistas ao aperfeiçoamento e acompanhamento da qualidade de ensino oferecido na UFSM. Os resultados desta avaliação serão utilizados no planejamento e na implementação de ações que busquem a melhoria da prática pedagógica na Instituição.	
Instruções do Questionário Prezado(a) Estudante, Este instrumento é constituído por um conjunto de questões e tem o propósito específico de avaliar o docente em relação a sua atuação na disciplina ministrada. Salieta-se que o instrumento deverá ser respondido com responsabilidade e seriedade, de modo que sua opinião possa contribuir com a promoção da qualidade do ensino na Instituição.	
1 - Avaliação do Docente pelo Discente	
Instruções da Seção Responda ao questionário utilizando a escala a seguir: (1) Discordo totalmente (se você discorda em 100% da afirmativa); (2) Discordo em parte (se você discorda da afirmativa, mas não em 100%); (3) Não concordo nem discordo (se você está indeciso ou neutro em relação a afirmativa); (4) Concordo em parte (se você concorda com a afirmativa, mas não em 100%); (5) Concordo totalmente (se você concorda em 100% da afirmativa); (6) Não sei responder (se você não tem conhecimento acerca da afirmativa).	
1.1 - O programa da disciplina foi apresentado pelo professor (objetivos, conteúdo a ser desenvolvido e bibliografia).*	
<input type="radio"/> Discordo totalmente <input type="radio"/> Discordo em parte <input type="radio"/> Não concordo nem discordo <input type="radio"/> Concordo em parte <input type="radio"/> Concordo totalmente <input type="radio"/> Não sei responder	
1.2 - O professor comparece às aulas e cumpre os horários de início e de término das mesmas.*	
<input type="radio"/> Discordo totalmente <input type="radio"/> Discordo em parte <input type="radio"/> Não concordo nem discordo <input type="radio"/> Concordo em parte <input type="radio"/> Concordo totalmente <input type="radio"/> Não sei responder	
1.3 - O professor utiliza de forma adequada os recursos didáticos disponíveis, favorecendo a aprendizagem.*	
<input type="radio"/> Discordo totalmente <input type="radio"/> Discordo em parte <input type="radio"/> Não concordo nem discordo <input type="radio"/> Concordo em parte <input type="radio"/> Concordo totalmente <input type="radio"/> Não sei responder	
1.4 - O professor estimula a utilização de materiais complementares (livros, sites, periódicos on line, áudio, vídeos, entre outros).*	
<input type="radio"/> Discordo totalmente <input type="radio"/> Discordo em parte <input type="radio"/> Não concordo nem discordo <input type="radio"/> Concordo em parte <input type="radio"/> Concordo totalmente <input type="radio"/> Não sei responder	
1.5 - O professor demonstra domínio sobre o conteúdo apresentado, tratando-o com clareza e objetividade.*	
<input type="radio"/> Discordo totalmente <input type="radio"/> Discordo em parte <input type="radio"/> Não concordo nem discordo <input type="radio"/> Concordo em parte <input type="radio"/> Concordo totalmente <input type="radio"/> Não sei responder	

1.6 - O professor estabelece relações entre os conteúdos da sua disciplina com os conteúdos das demais disciplinas, contribuindo com a formação profissional.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.7 - O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.8 - O professor mostra-se disponível para tirar dúvidas sobre a disciplina e o conteúdo ministrado.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.9 - O professor ouve críticas, opiniões e sugestões referentes às suas aulas, mostrando-se aberto ao diálogo.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.10 - O professor apresenta uma postura de respeito mútuo, ou seja, preserva a imagem da Instituição, dos colegas, dos acadêmicos e não manifesta atitudes preconceituosas.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.11 - O professor elabora avaliações compatíveis com os conteúdos ministrados em aula.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.12 - Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.13 - O professor cumpre o programa da disciplina apresentado.*

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.14 - O professor tem aluno de pós-graduação em docência orientada que ministra aulas teóricas ou práticas para a turma?*

Sim Não

1.15 - O professor acompanha o aluno de pós-graduação durante a docência orientada.

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

1.16 - O professor cumpre a Resolução UFSM N. 018/2008, que estabelece que seja ministrada no máximo 30% da carga horária da disciplina em docência orientada.

Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

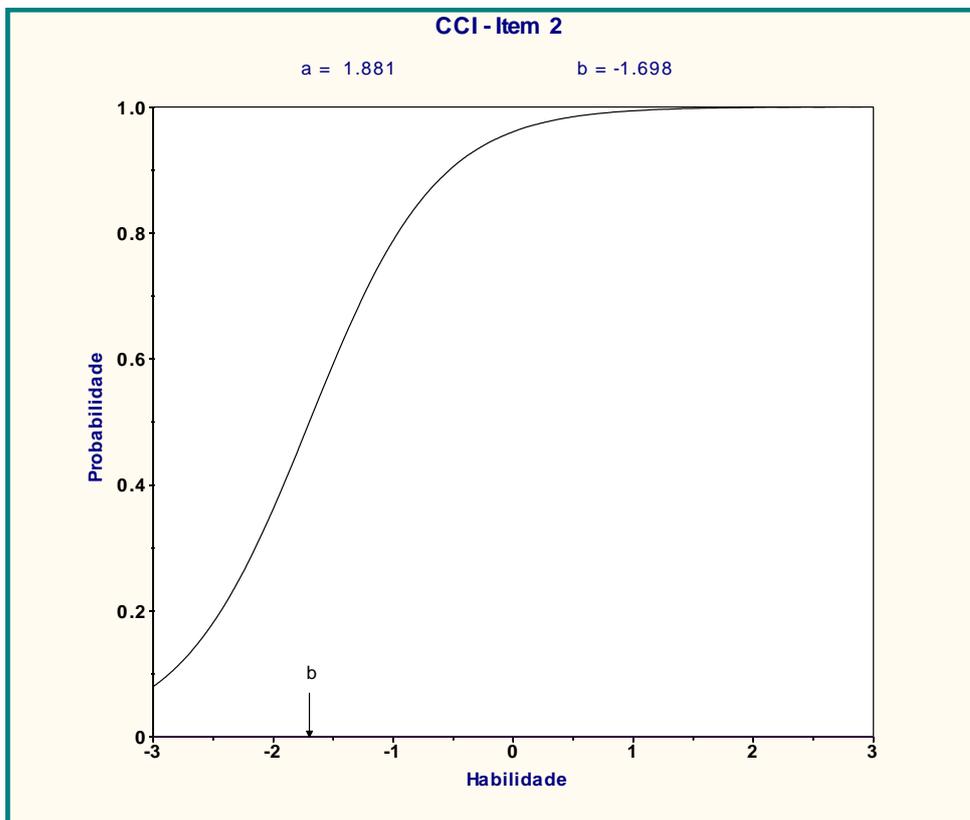
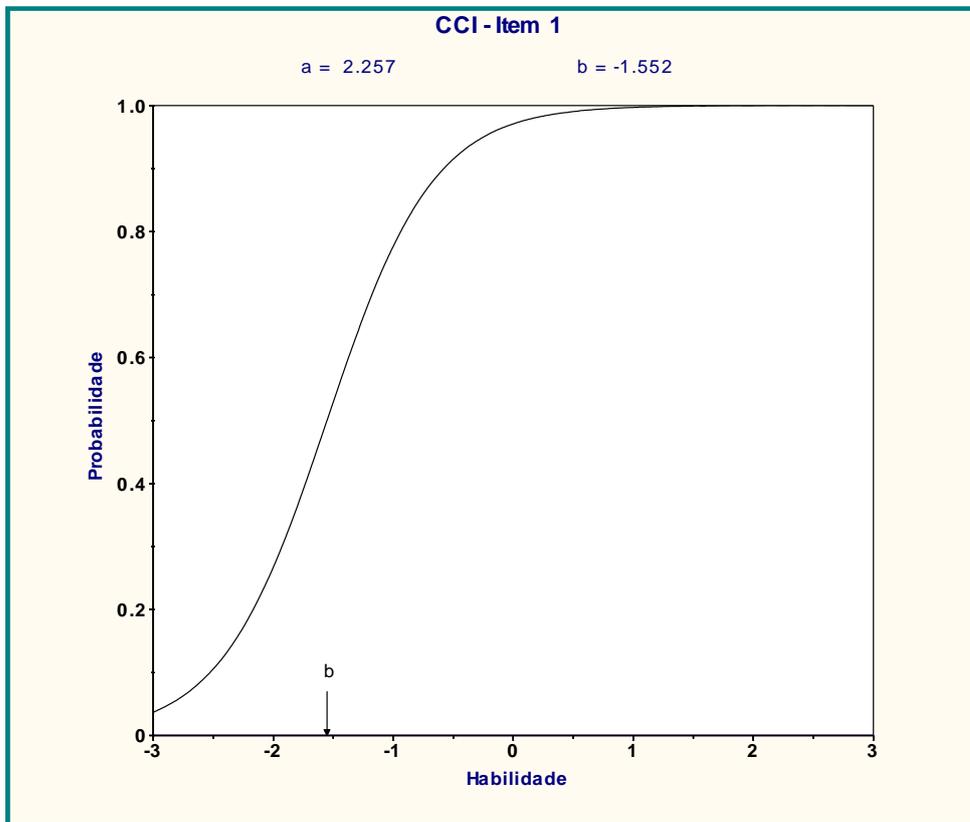
1.17 - Como discente, eu me comprometi com as atividades propostas pelo professor que acabei de avaliar (fui assíduo nas aulas, respeitei os horários de aula chegando e saindo nos horários previstos, dediquei-me aos exercícios, trabalhos, provas e debates propostos pelo professor em aula).*

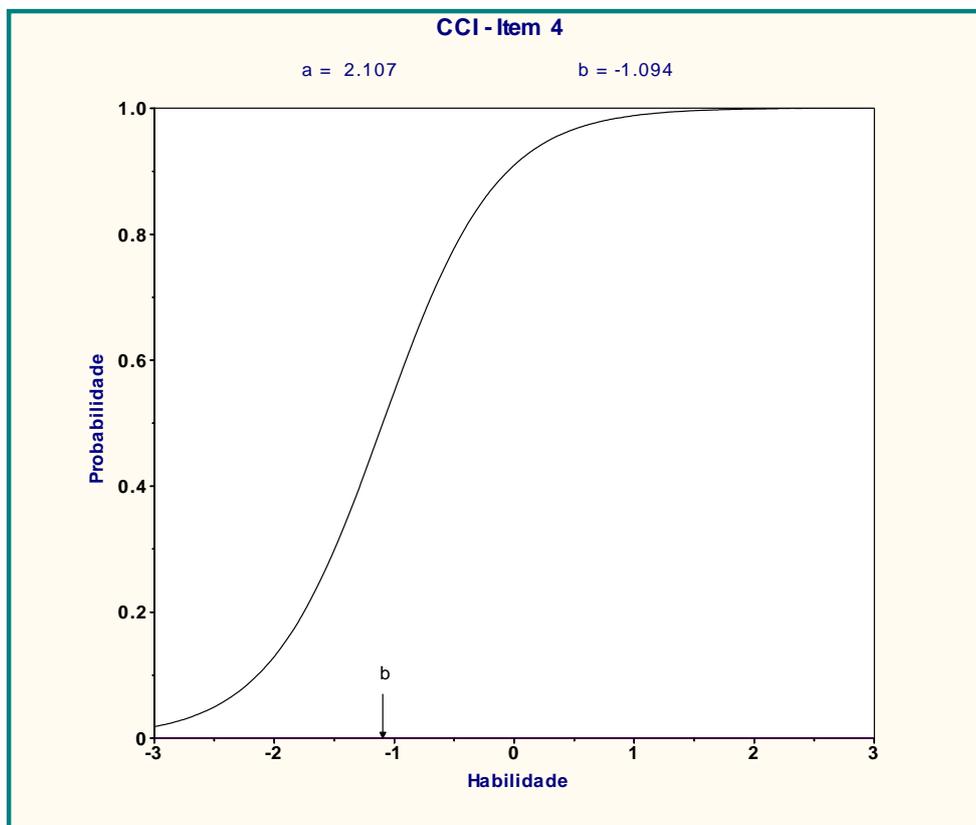
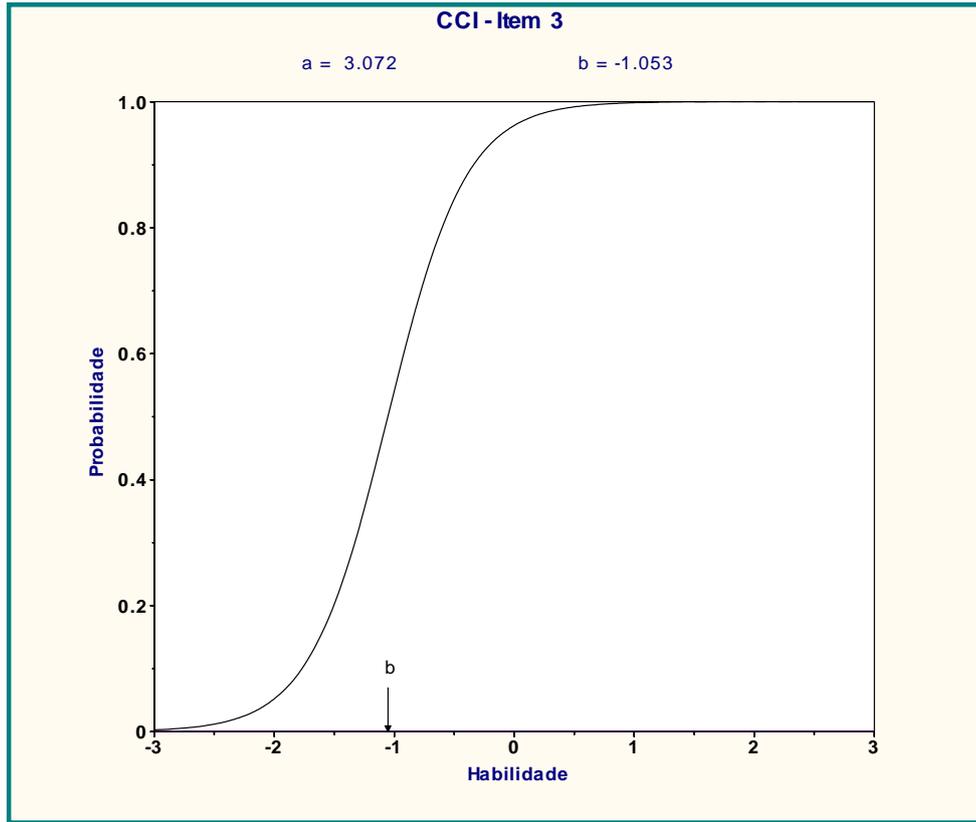
Discordo totalmente Discordo em parte Não concordo nem discordo Concordo em parte Concordo totalmente Não sei responder

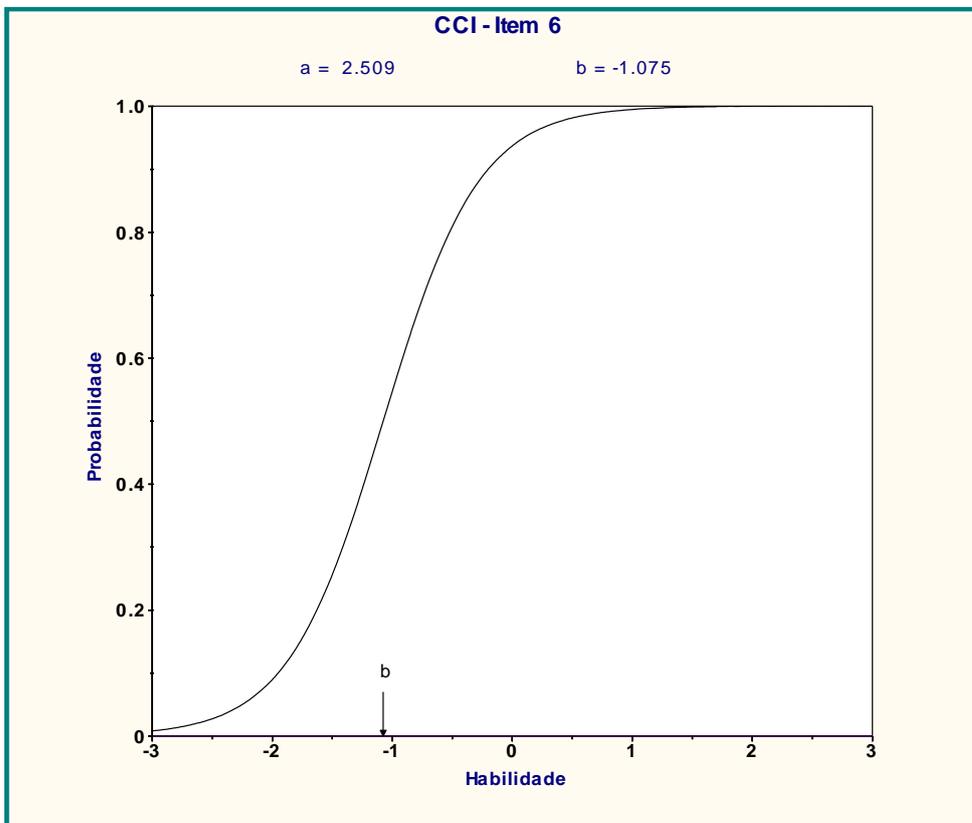
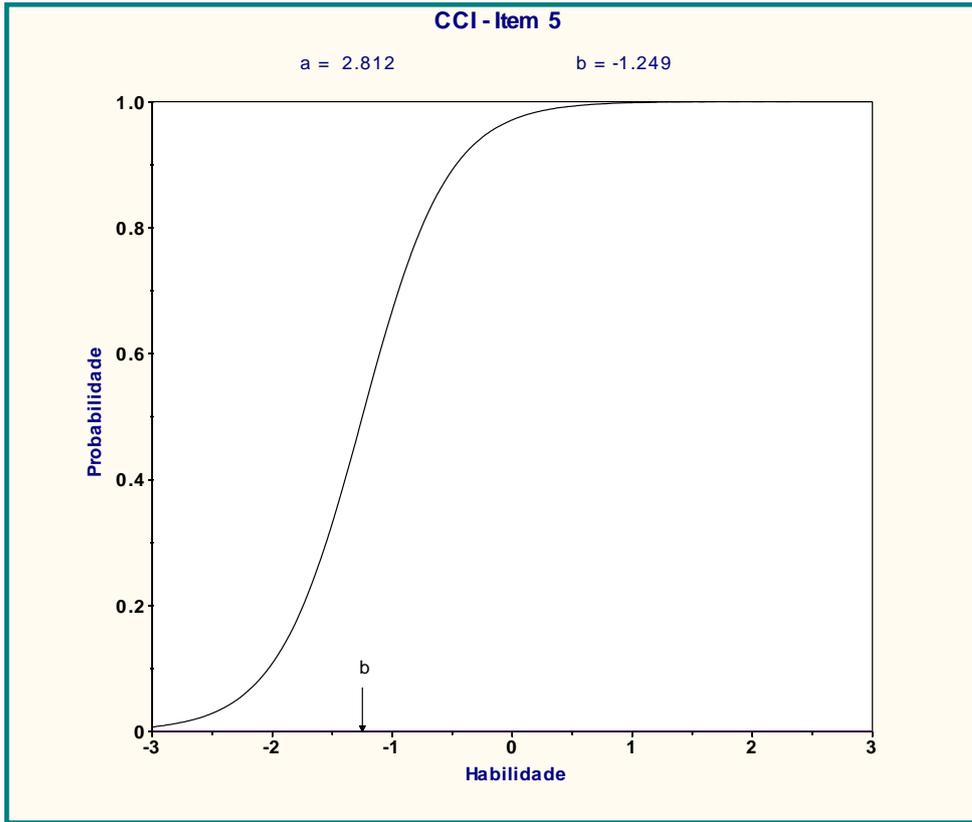
1.18 - Espaço destinado a sugestões, opiniões ou críticas.

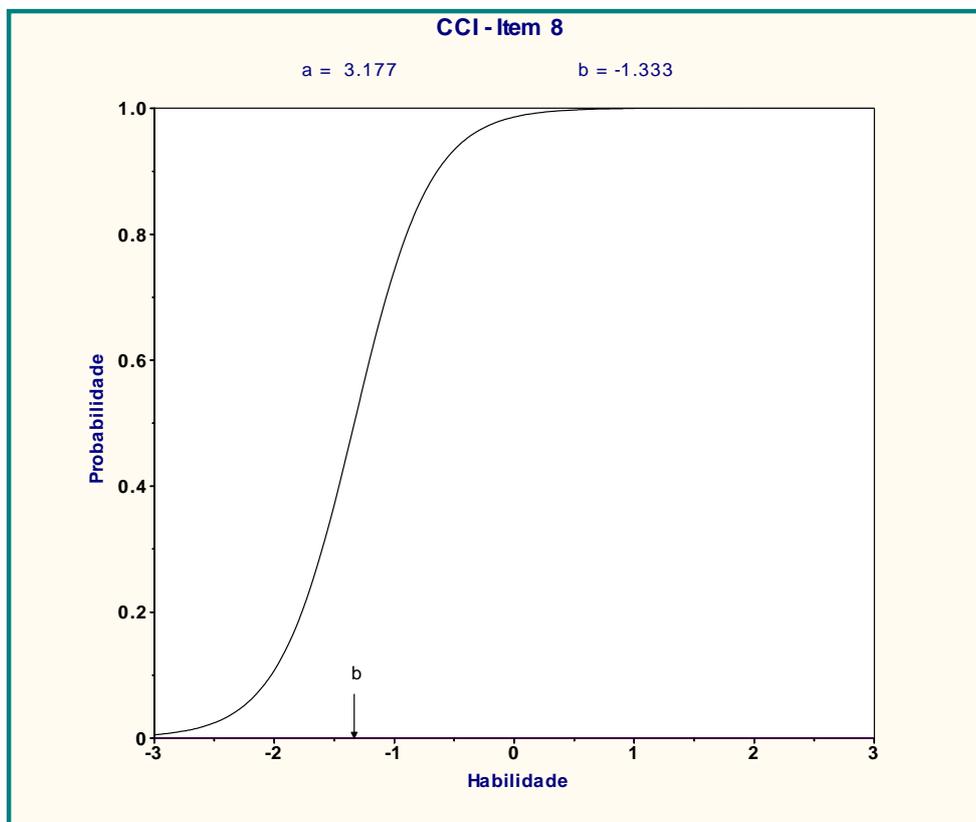
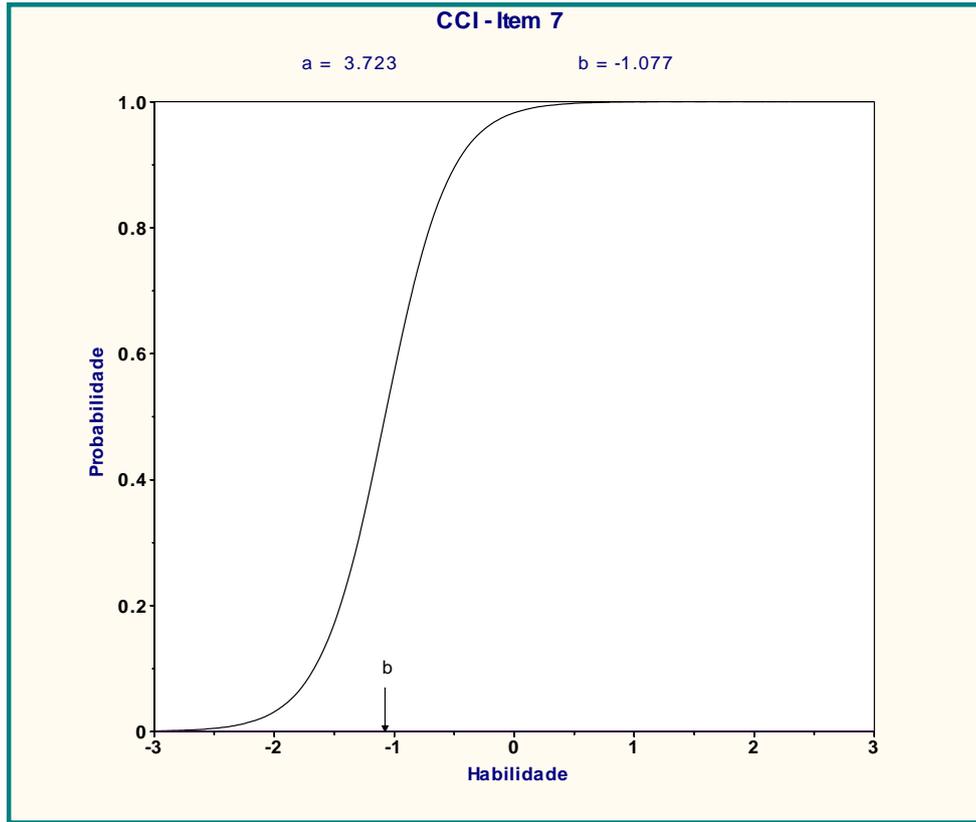
Restam 5120 caracteres | Caracteres: 0 | Palavras: 0

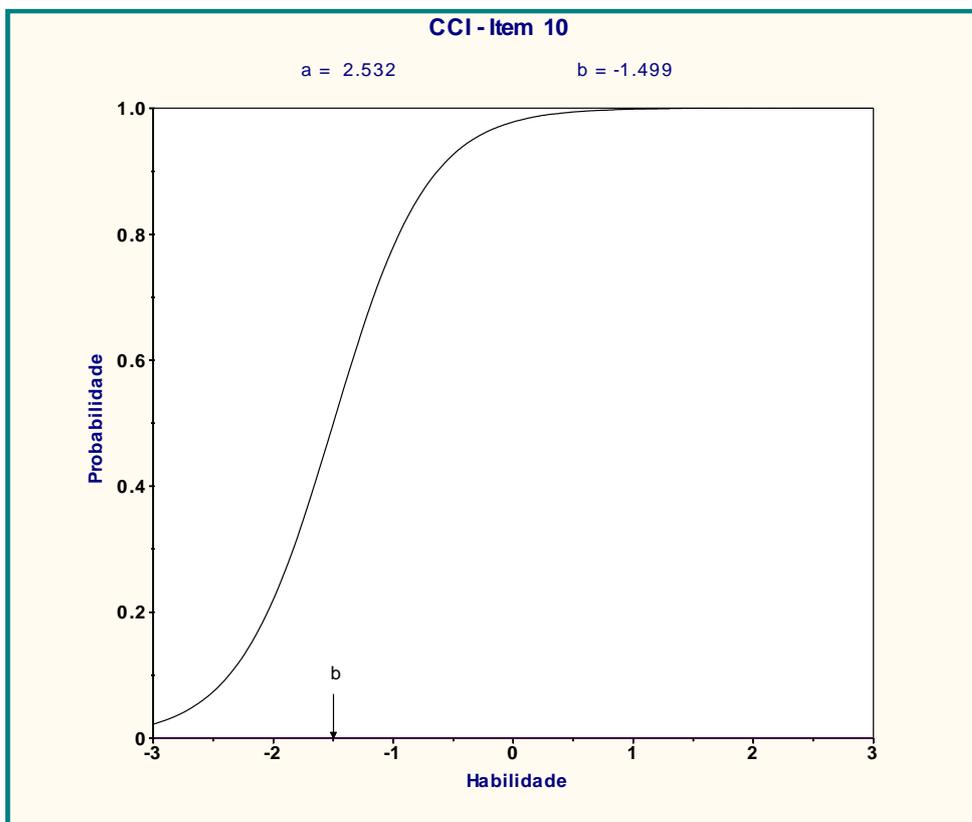
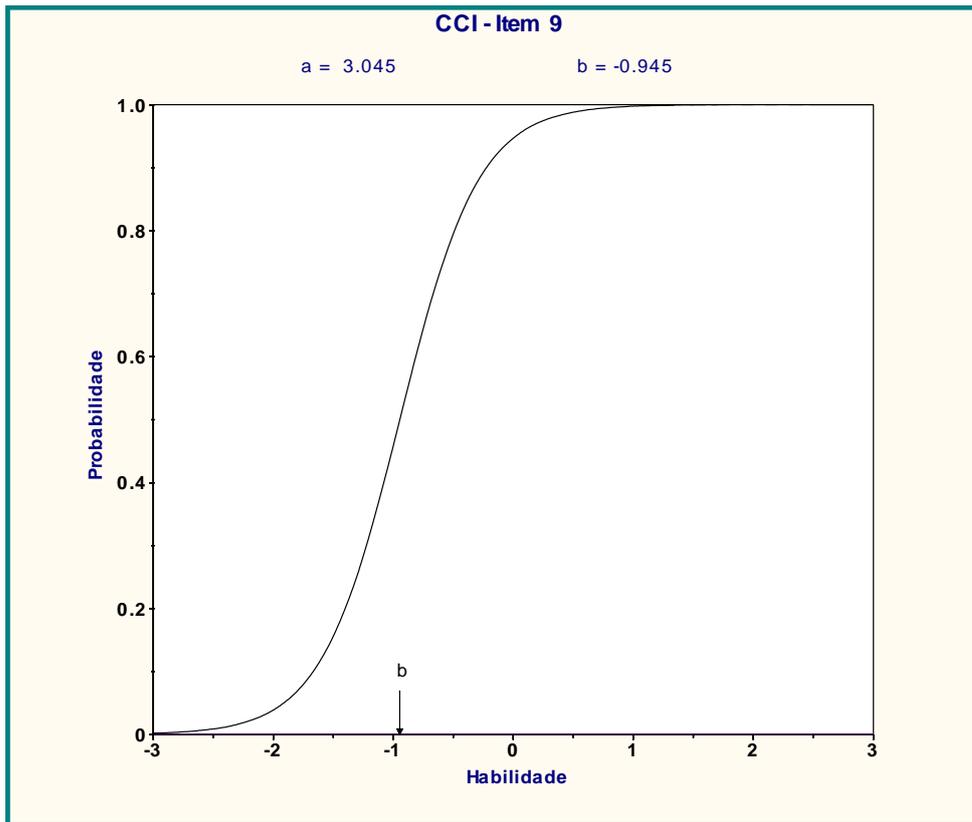
ANEXO B – Curvas Características dos Itens (CCI)

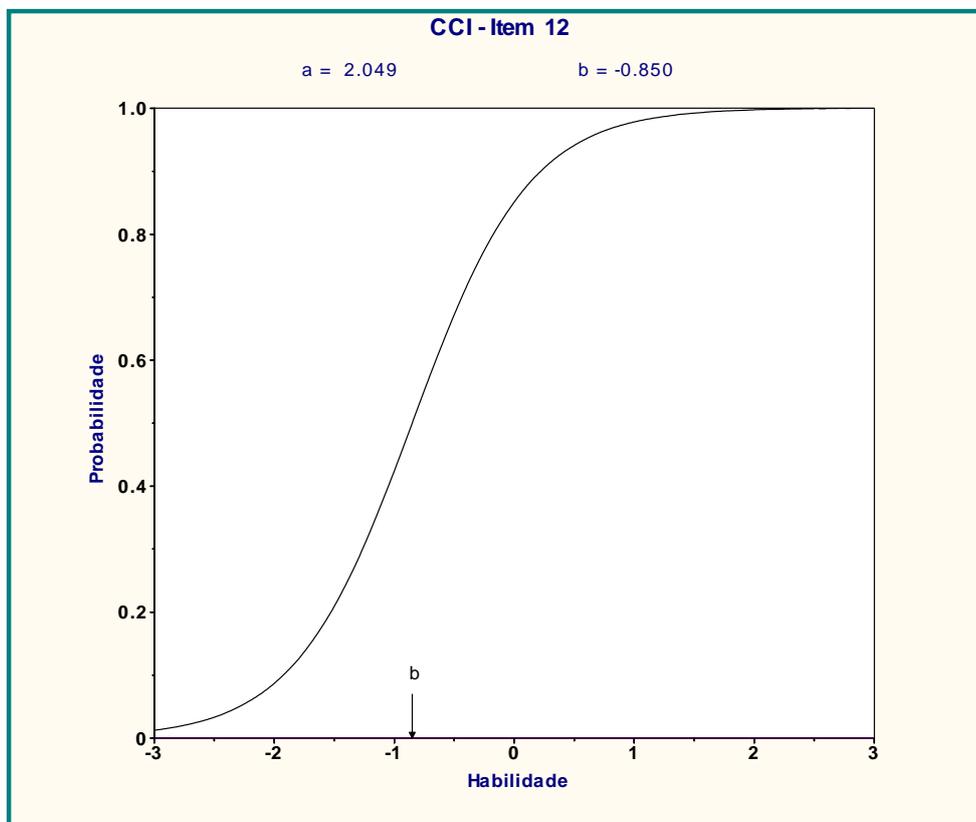
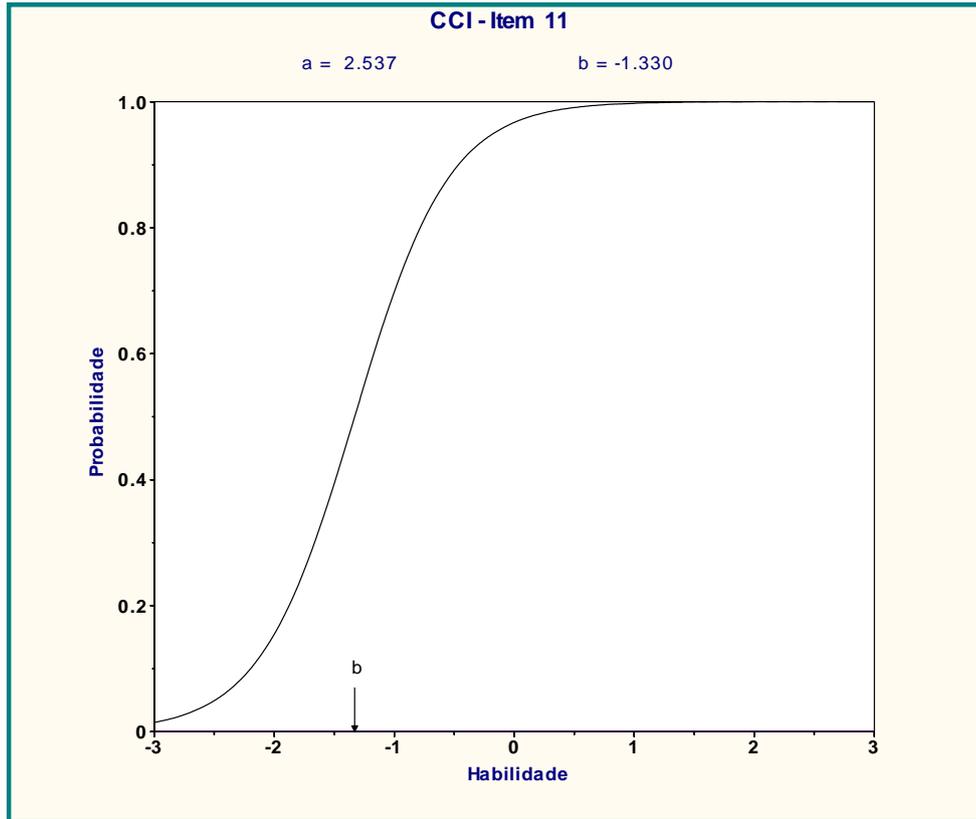


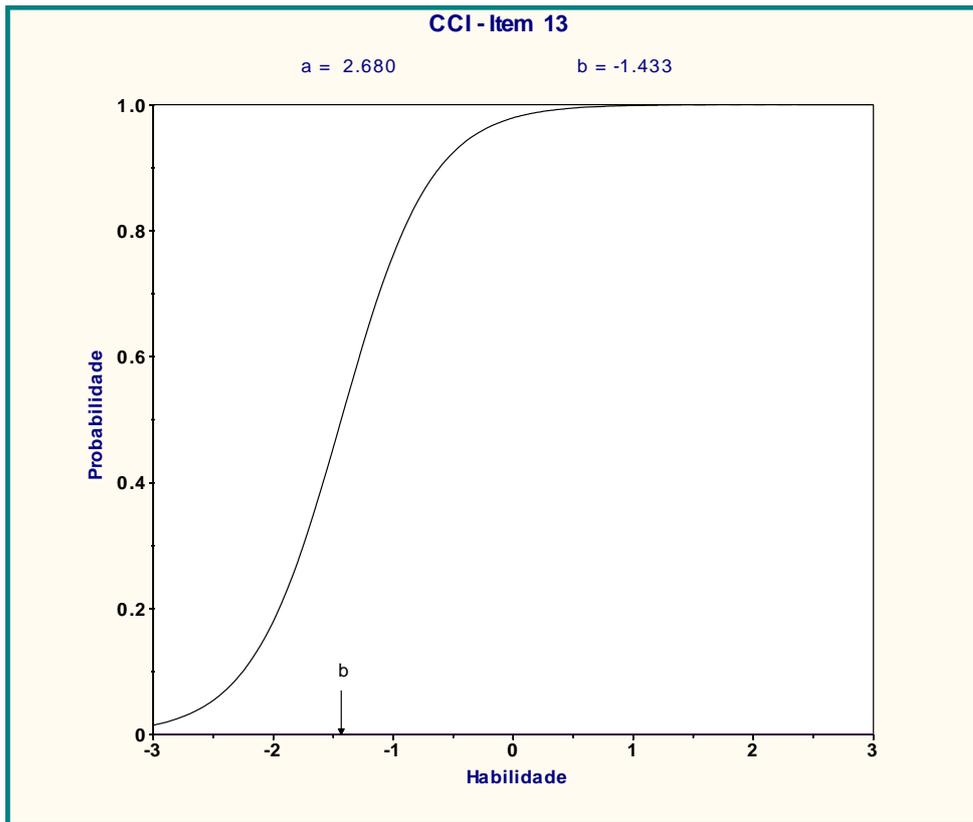












ANEXO C – Curvas de Informação dos Itens (CII)

