

Thiago Korb

**A AVALIAÇÃO DO DOMÍNIO PSICOMOTOR DA
APRENDIZAGEM: ESTUDO PARA CONSTRUÇÃO DE
ESCALA DE PROFICIÊNCIA NA EDUCAÇÃO PROFISSIONAL**

Dissertação submetida ao Programa de Pós-Graduação em Métodos e Gestão em Avaliação (PPGMGA) da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Métodos e Gestão em Avaliação.

Orientador: Prof. Dr. Dalton Francisco de Andrade

Florianópolis
2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Korb, Thiago

A Avaliação do domínio psicomotor da aprendizagem :
Estudo para construção de escala de proficiência na educação
profissional / Thiago Korb ; orientador, Dalton Francisco
de Andrade, 2019.
169 p.

Dissertação (mestrado profissional) - Universidade
Federal de Santa Catarina, Centro Tecnológico, Programa de
Pós-Graduação em Métodos e Gestão em Avaliação, Florianópolis,
2019.

Inclui referências.

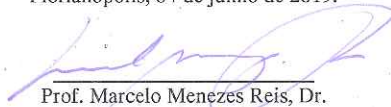
1. Métodos e Gestão em Avaliação. 2. Domínio Psicomotor.
3. Educação Profissional. 4. Teoria de Resposta ao Item. I.
Andrade, Dalton Francisco de . II. Universidade Federal de
Santa Catarina. Programa de Pós-Graduação em Métodos e
Gestão em Avaliação. III. Título.

Thiago Korb

**A AVALIAÇÃO DO DOMÍNIO PSICOMOTOR DA APRENDIZAGEM:
ESTUDO PARA CONSTRUÇÃO DE ESCALA DE PROFICIÊNCIA NA
EDUCAÇÃO PROFISSIONAL**

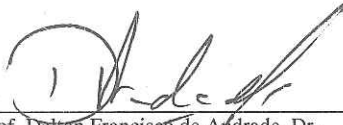
Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre em Métodos e Gestão em Avaliação” e aprovada em sua forma final pelo Programa de Pós-Graduação em Métodos e Gestão em Avaliação.

Florianópolis, 04 de junho de 2019.



Prof. Marcelo Menezes Reis, Dr.
Coordenador do Curso

Banca Examinadora:



Prof. Dalton Francisco de Andrade, Dr.
Orientador

Universidade Federal de Santa Catarina (UFSC)



Prof. Carlos Henrique Sancineto da Silva Nunes, Dr.
Universidade Federal de Santa Catarina (UFSC)



Prof.ª Ligia Maria Vettorato Trevisan, Dra.
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)



Juliano Anderson Pacheco, Dr.
Federação das Indústrias do Estado de Santa Catarina (FIESC)

Este trabalho é dedicado à minha amada mãe e a meu querido pai (*in memoriam*).

AGRADECIMENTOS

Primeiramente gostaria de agradecer a Deus, que em sua infinita e magnífica bondade, me deu forças para lutar e conquistar esse sonho.

A minha família, Valmira, Camila, Tereza, Zeno e Fernando, que em diversos momentos me deram luz, amparo e carinho, ouvindo e respeitando minhas necessidades de estudos. A vocês que souberam esperar meu tempo e que torcem pelo meu sucesso, minha mais profunda gratidão.

Ao meu orientador, prof. Dr. Dalton Francisco de Andrade, por me receber inúmeras vezes e compartilhar seus ensinamentos. Você é um exemplo como educador em que buscarei me espelhar e espero dedicar-me com o mesmo afinho aos meus alunos.

Ao SENAI, instituição que atuo há mais de dez anos, pela oportunidade e confiança pelo trabalho até aqui desenvolvido. Certamente o apoio de todos colegas, em especial a equipe do Departamento Nacional, Hugo, Gleci e Rosa, foi fundamental para que esta pesquisa se concretizasse. Ao Departamento Regional de SC, meu agradecimento pelo tempo disponibilizado para minha formação.

Aos meus colegas de trabalho e, também, amigos, Cleunisse, Regina, Tiago Conte, Fernanda, Maristela, Myrthes, Rangélia e Ana Fontes, muito obrigado por me ouvir e incentivar nessa conquista. Vocês acompanharam, ainda que na parcialidade, muitas das descobertas dessa trajetória.

Aos amigos de vida e de longa data, aqueles que chegaram e ficaram por pura afinidade, Sheila, Márcia, Carol e Jéssica, gratidão pela compreensão em não estar tão presente nesses últimos meses. Vocês me conhecem, sabem do quanto as aprecio e que sempre as carregarei no coração.

Aos meus colegas de jornada nessa missão do mestrado, em destaque a Maristeele, Alexandra, Bianca, Cássio e Deise, que partilharam muitos momentos de auxílio mútuo e de companheirismo.

Aos membros da banca examinadora, Prof. Dr. Carlos Henrique Sancineto da Silva Nunes, Prof^a. Dra. Ligia Maria Vettorato Trevisan e

Dr. Juliano Anderson Pacheco, pelas excelentes contribuições indicadas e que certamente abrilhantaram ainda mais essa pesquisa. Agradeço a disponibilidade, cordialidade e apontamentos pertinentes.

A todos vocês, o meu sincero agradecimento!

“Aquele que quer aprender a voar um dia precisa primeiro aprender a ficar de pé, caminhar, correr, escalar e dançar; ninguém consegue voar só aprendendo voo.”

(Friedrich Nietzsche)

RESUMO

A análise da proficiência de estudantes é um dos principais componentes em qualquer sistema de avaliação em larga escala, porém apenas a dimensão cognitiva é predominantemente observada, além da aplicação generalizada de testes no modelo papel caneta. A educação profissional precisa ir além, pois mais do que “saber”, o estudante precisa “fazer”, ou seja, o importante é conseguir desenvolver na prática suas competências profissionais. Nessa perspectiva, o objetivo deste trabalho é propor uma escala que permita avaliar na prática o domínio psicomotor de estudantes da educação profissional à luz da Teoria de Resposta ao Item (TRI). A população era composta por 5.856 estudantes, do Curso Técnico em Mecânica, do Serviço Nacional de Aprendizagem Industrial (SENAI), sendo selecionada amostra de 1.498 representantes (25,6%) para testagem. Destaca-se que a coleta de dados foi de responsabilidade da instituição, restringindo-se a análise dos dados e proposição de um método para interpretação. As principais metodologias empregadas foram a Revisão Sistemática da Literatura (RSL) para mapeamento de experiências anteriores, aplicação do Modelo Logístico de 2 Parâmetros (ML2P) para calibração dos itens e estimação das proficiências, e definição de critérios estatísticos e técnicos para construção da escala de desempenho. Os resultados apresentaram boa aderência ao modelo TRI utilizado, sendo descartado apenas um dos 32 itens avaliados. A Curva de Informação do Teste (CIT) identificou que o instrumento apresenta informação mais relevante para habilidades no intervalo -2,2 a -0,4 na escala (0, 1), portanto, uma prova considerada de fácil para médio. As evidências foram posicionadas conforme os níveis âncora definidos e segmentas em três cortes, sendo que 15,55% da amostra foi classificada como “elementar”, 52,47% pertencem ao nível “moderado” e 31,98% a categoria “pleno”. Foram sete as principais contribuições concluídas com esse estudo: (1) realce do domínio psicomotor como elemento fundamental na avaliação de estudantes da educação profissional; (2) a RSL identificou poucas publicações mapeadas que tratam desse aspecto, demonstrando a escassez de instrumentos práticos em aplicações dessa natureza; (3) construção do constructo orientada pela taxonomia dos objetivos educacionais adaptada de Dave (1970) e Simpson (1972), bem como a Pirâmide de Miller (1990); (4) êxito na aplicação da TRI como metodologia de análise dos resultados; (5) caráter unidimensional do constructo, embora chama-se atenção para estudos futuros sobre a bidimensionalidade; (6) a possibilidade de realizar a interpretação

pedagógica de modo sumário ou detalhado, conforme a necessidade e público; e (7) apontamento das diferenças entre as proficiências dos estudantes quando comparados por Unidade Federativa (UF), sexo e idade.

Palavras-chave: Domínio Psicomotor. Educação Profissional. Teoria de Resposta ao Item.

ABSTRACT

Student proficiency analysis is one of the main components of any large-scale evaluation system, but only the cognitive dimension is predominantly observed, as well as the generalized application of tests in the pen-paper model. Professional education needs to go beyond, because more than "to know", the students needs "how to do", that is, the important thing is to be able to develop their professional skills in practice. In this perspective, the objective of this work is to propose a scale that allows to evaluate in practice the psychomotor domain of students of professional education in the light of the Item Response Theory (IRT). The population was composed of 5,856 students, from the Technical Course in Mechanics, of the National Service of Industrial Training (SENAI), being selected a sample of 1,498 representatives (25.6%) for testing. It stands out that the data collection was the responsibility of the institution, restricting the data analysis and proposing a method for interpretation. The main methodologies used were the Systematic Review for mapping previous experiences, application of the Two-Parameter Logistic Model (2PL) for item calibration and proficiency estimation, and definition of statistical and technical criterion for construction the performance scale. The results showed good adherence to the IRT model used, being discarded only one of the 32 items evaluated. The Test Characteristic Curve (TCC) identified that the instrument presents information more relevant to abilities in the range -2.2 to -0.4 on the scale (0, 1), thus a test considered easy to medium. The evidences were positioned according to the defined anchorage levels and classified in three cuts, with 15.55% of the sample classified as "elementary", 52.47% belong to the "moderate" level and 31.98% to the "full". There were seven main contributions completed with this study: (1) highlight of the psychomotor domain as a fundamental element in the evaluation of professional education students; (2) the RSL has identified few mapped publications dealing with this aspect, demonstrating the scarcity of practical tools in such applications; (3) construction of the construct guided by the Taxonomy of Educational Objectives of Dave (1970) and Simpson (1972), as well as the Miller's Pyramid (1990); (4) successful implementation of IRT as a methodology for analyzing results; (5) one-dimensional character of the construct, although attention is drawn to future studies on two-dimensional; (6) the possibility of performing the pedagogical interpretation in a summary or detailed manner, according to the need and public; and (7) pointing out the

differences between students' proficiencies when compared to the federative unit (UF), gender and age.

Keywords: Psychomotor Domain. Professional education. Item Response Theory.

LISTA DE FIGURAS

Figura 1 - Componentes da competência	45
Figura 2 - Categorias da dimensão cognitiva (Taxonomia de Bloom)..	48
Figura 3 - Categorias da dimensão cognitiva (Taxonomia de Bloom Revisada).....	49
Figura 4 - Categorias da dimensão psicomotora (Taxonomia de Bloom Revisada).....	50
Figura 5 - Categorias da dimensão afetiva (Taxonomia de Bloom Revisada).....	50
Figura 6 - Pirâmide de Miller.....	51
Figura 7 - As características das rubricas (evidências)	59
Figura 8 – Escolha de modelos de TRI	69
Figura 9 - Fases da metodologia de pesquisa	87
Figura 10 - Matriz de referência do Curso Técnico em Mecânica	96
Figura 11 - Captura de tela da Plataforma SAP (exemplo fictício).....	105
Figura 12 - Posicionamento das evidências com especialistas.....	110
Figura 13 - Distribuição das proficiências dos estudantes avaliados ..	125
Figura 14 - Posicionamento das evidências pelo critério estatístico ...	132
Figura 15 - Posicionamento das evidências na escala de desempenho	133
Figura 16 - Pontos de corte dos níveis de desempenho.....	135

LISTA DE QUADROS

Quadro 1 - Comparação entre a TCT e TRI.....	64
Quadro 2 - Experiências mapeadas na RSL	77
Quadro 3 - Filtros aplicados na RSL	90
Quadro 4 - Correlação bisserial das evidências avaliadas.....	118
Quadro 5 - Classificação das evidências quanto ao parâmetro de discriminação	120
Quadro 6 - Descrição pedagógica sumária de cada nível de desempenho	137
Quadro 7 - Correlação das capacidades do domínio cognitivo (matriz de referência) x domínio psicomotor (nova proposta)	138
Quadro 8 - Classificação das capacidades psicomotoras.....	139
Quadro 9 - Descrição pedagógica detalhada dos níveis de desempenho	140

LISTA DE TABELAS

Tabela 1 - Distribuição da amostra de estudantes por UF e Sexo	115
Tabela 2 - Distribuição da amostra de estudantes por UF e Idade	116
Tabela 3 - Proporções de erros/acertos por evidência avaliada.....	117
Tabela 4 - Calibração dos parâmetros “a” e “b” das evidências.....	119
Tabela 5 - Cargas fatoriais comparando modelo unidimensional e bidimensional	129
Tabela 6 - Distribuição dos níveis de desempenho da escala por UF .	143
Tabela 7 - Distribuição dos níveis de desempenho da escala por sexo	144
Tabela 8 - Distribuição dos níveis de desempenho da escala por classes de idade	144

LISTA DE GRÁFICOS

Gráfico 1 - Matrículas de Educação Profissional Técnica	42
Gráfico 2 - Curva Característica do Item (CCI)	70
Gráfico 3 - Distribuição da amostra de estudantes por UF.....	114
Gráfico 4 - Curvas características das evidências E06 e E22.....	122
Gráfico 5 - Curvas de informação das evidências E07 e E08	123
Gráfico 6 - CIT da prova prática do Curso Técnico em Mecânica.....	124
Gráfico 7 - Distribuição das proficiências estimadas	126
Gráfico 8 - Análise da dimensionalidade do instrumento	128
Gráfico 9 - Valor de α conforme eliminação de cada evidência	130
Gráfico 10 - CIT comparada com os níveis de desempenho da escala	136
Gráfico 11 - Distribuição da amostra por níveis de desempenho.....	142

LISTA DE ABREVIATURAS E SIGLAS

ACOS – Colégio Americano de Cirurgiões Osteopáticos
AM – Amazonas
BA – Bahia
Biss. – Bisserial
CBO – Classificação Brasileira de Ocupação
CCI – Curva Característica do Item
CE – Ceará
CEFET – Centros de Formação de Educação Profissionais
CII – Curva de Informação do Item
CIT – Curva de Informação do Teste
CK – Conhecimento de conteúdo
CNCT – Catálogo Nacional de Cursos Técnicos
EAP – Esperança a posteriori
E.P. – Erro Padrão
EPI – Equipamentos de Proteção Individual
ES – Espírito Santo
Ev. – Evidência
FIT – Função de Informação do Teste
FRI – Função de Resposta do Item
GO – Goiás
INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio
Teixeira
LDB – Lei de Diretrizes e Bases da Educação Brasileira
MAP – Moda a posteriori
MCP – Modelo de crédito parcial
MCPG – Modelo de crédito parcial generalizado
MEC – Ministério da Educação
MEG – Modelo de escala gradual
MG – Minas Gerais
ML1P – Modelo logístico de um parâmetro
ML2P – Modelo logístico de dois parâmetros
ML3P – Modelo logístico de três parâmetros
MRG – Modelo de respostas graduais
MRN – Modelo de respostas nominais
MS – Mato Grosso do Sul
MSEP – Metodologia SENAI de Educação Profissional
MT – Mato Grosso
MV – Máxima Verossimilhança
MVM – Máxima Verossimilhança Marginal

NBOME – Conselho Nacional de Examinadores Médicos Osteopatas
PCK – conhecimento pedagógico de conteúdo
PE – Pernambuco
PNE – Plano Nacional de Educação
PR – Paraná
PROEP – Programa de Expansão da Educação Profissional
Pronatec – Programa Nacional de Acesso ao Ensino Técnico e Emprego
RJ – Rio de Janeiro
RS – Rio Grande do Sul
RSL – Revisão Sistemática da Literatura
SAEB – Sistema de Avaliação da Educação Básica
SAEP – Sistema de Avaliação da Educação Profissional
SAP – Sistema de Avaliação Prática
SC – Santa Catarina
SE – Sergipe
SENAC – Serviço Nacional de Aprendizagem Comercial
SENAI – Serviço Nacional de Aprendizagem Industrial
SP – São Paulo
TCM – Teoria Clássica das Medidas
TCT – Teoria Clássica do Teste
TO – Tocantins
TRI – Teoria de Resposta ao Item
UF – Unidades Federativas
UFRGS – Universidade Federal do Rio Grande do Sul
UFSC – Universidade Federal de Santa Catarina
UO – Unidade Operacional

SUMÁRIO

1 INTRODUÇÃO.....	29
1.1 OBJETIVOS.....	30
1.1.1 Objetivo geral.....	31
1.1.2 Objetivos específicos.....	31
1.2 DEFINIÇÃO DO PROBLEMA.....	31
1.3 JUSTIFICATIVA DO PROBLEMA.....	31
1.4 DELIMITAÇÃO DO TEMA.....	32
1.5 ESTRUTURA DA DISSERTAÇÃO.....	33
2 FUNDAMENTAÇÃO TEÓRICA.....	35
2.1 EDUCAÇÃO PROFISSIONAL NO BRASIL.....	35
2.1.1 Histórico da educação profissional no Brasil.....	36
2.1.2 A educação profissional de nível técnico.....	41
2.2 AVALIAÇÃO DO DOMÍNIO PSICOMOTOR DA APRENDIZAGEM.....	43
2.2.1 Avaliação por competências.....	45
2.2.2 Domínios da aprendizagem.....	47
2.2.3 Matriz de referência para avaliações em larga escala.....	52
2.2.4 Instrumentos de avaliação.....	54
2.2.5 Avaliar por meio de evidências.....	58
2.3 TEORIA DE RESPOSTA AO ITEM.....	60
2.3.1 TRI x TCT.....	62
2.3.2 Pressupostos epistemológico da TRI.....	65
2.3.3 Modelos de TRI.....	66
2.3.4 Construção da escala do traço latente.....	71
2.3.5 Validade e fidedignidade do teste.....	75
2.3.6 TRI aplicada à avaliação da educação profissional.....	77
3 PROCEDIMENTOS METODOLÓGICOS.....	85
3.1 CARACTERIZAÇÃO DA PESQUISA.....	85
3.2 REVISÃO SISTEMÁTICA DA LITERATURA.....	89
3.3 POPULAÇÃO E AMOSTRA DA PESQUISA.....	91

3.4	CONSTRUCTO DO DOMÍNIO PSICOMOTOR	93
3.5	INSTRUMENTO DE AVALIAÇÃO	97
3.5.1	Caderno de prova	98
3.5.2	Lista de verificação	99
3.5.3	Manual do Avaliador.....	100
3.5.4	Lista de verificação	100
3.6	PRÉ-TESTE	101
3.7	APLICAÇÃO DAS PROVAS	102
3.7.1	Planejamento	102
3.7.2	Execução	104
3.7.3	Registros da aplicação.....	105
3.8	TRATAMENTO DOS DADOS.....	106
3.9	CONSTRUÇÃO, INTERPRETAÇÃO E APLICAÇÃO DA ESCALA.....	110
4	ANÁLISE E DISCUSSÃO DOS RESULTADOS.....	113
4.1	ANÁLISE ESTATÍSTICA DESCRITIVA.....	113
4.2	ANÁLISE E CALIBRAÇÃO DAS EVIDÊNCIAS.....	116
4.2.1	Análise pela TCT	116
4.2.2	Calibração pela TRI	118
4.2.3	Análise gráfica das evidências	121
4.3	ESTIMAÇÃO DAS PROFICIÊNCIAS.....	124
4.4	ANÁLISE DO INSTRUMENTO DE AVALIAÇÃO	126
4.5	CONSTRUÇÃO DA ESCALA	131
4.6	INTERPRETAÇÃO PEDAGÓGICA	134
4.7	APLICAÇÃO DA ESCALA DE DESEMPENHO.....	142
5	CONSIDERAÇÕES FINAIS.....	147
	REFERÊNCIAS.....	153
	APÊNDICES	170
	APÊNDICE A - Script R: Dimensionalidade do Teste.....	170
	APÊNDICE B - Sintaxe Bilog-MG	171

APÊNDICE C - Curvas características das evidências e curvas de informação das evidências	172
APÊNDICE D - Posicionamento das evidências na escala de proficiência.....	181
ANEXOS	182
ANEXO A - Caderno de prova	182
ANEXO B - Lista de infraestrutura.....	183
ANEXO C - Manual do avaliador	184
ANEXO D - Lista de verificação	185
ANEXO E - Termo de autorização para realização da pesquisa.....	191
ANEXO F - Base de dados	192

1 INTRODUÇÃO

Atualmente, no Brasil, não há um sistema de avaliação em larga escala para avaliação da Educação Profissional, a exemplo do Sistema Nacional de Avaliação da Educação Superior (SINAES) e o Sistema de Avaliação da Educação Básica (SAEB), o que aponta uma lacuna e o não atendimento dos artigos 206, inciso VII que estabelece como princípio a “garantia do padrão de qualidade” e 209, inciso II “autorização e avaliação de qualidade pelo Poder Público”, respaldados pela lei de diretrizes e bases da educação nacional (LDB), Lei nº 9.394 de 20 de dezembro de 1996. Nos sistemas elencados, a análise do desempenho de proficiência dos estudantes é um dos principais componentes, configurando-se como um importante resultado e entendimento de que a educação profissional poderia seguir o mesmo caminho.

Além disso, o Decreto nº 7.690 de 2 de março de 2012, que estabelece as competências da Secretaria de Educação Profissional e Tecnológica (SETEC), do Ministério da Educação (MEC), corrobora com a perspectiva de que é necessário “estabelecer diretrizes para as ações de expansão e avaliação da Educação Profissional e Tecnológica em consonância com o Plano Nacional de Educação (PNE)”. Entretanto, não há uma política pública de avaliação de educação profissional no País, o que demanda reflexão e análise urgente frente à ampliação da oferta de vagas nessa modalidade, catalisada em especial pelo Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec), instituído pela Lei nº 12.513, de 26 de outubro de 2011. Em 2015, o MEC estruturou um grupo de trabalho para discutir uma política nessa vertente, porém o processo caminha com morosidade, sem previsão de implantação.

Diante dessa necessidade surge a demanda de criar uma metodologia que seja capaz de fazer essa avaliação. Entretanto, o modelo convencional por meio de testes cognitivos não supre todas as necessidades. A complexidade da educação profissional é maior, como destaca Francisco Soares, ex-presidente do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP): “as questões de um exame para estudantes, por exemplo, não podem ser lápis e papel. Uma das saídas é utilizar simuladores nas situações problema. O importante na educação profissional é o saber fazer” (CNI, 2015). Esta declaração indica a necessidade de testes práticos com estudantes, como foco em avaliar capacidades técnicas na prática.

Nessa vertente, esta pesquisa buscou apresentar uma proposta metodológica para interpretação de resultados de estudantes da educação profissional do Serviço Nacional de Aprendizagem Industrial (SENAI),

por meio da aplicação de provas práticas realizada com alunos do Curso Técnico em Mecânica. O SENAI é um dos cinco maiores complexos deste segmento no mundo e o maior da América Latina, apontado pela Organização das Nações Unidas (ONU) como uma das principais instituições educacionais do Hemisfério Sul. Fundada em 1942, já formou mais de 73,3 milhões de trabalhadores, em 28 áreas distintas de atendimento ao setor industrial.

O constructo a ser avaliado, em vista disso, foi o domínio psicomotor da aprendizagem, aportado no mapeamento da taxonomia dos objetivos educacionais de Bloom (1983) e, posteriormente, detalhada por outros autores como Dave (1970) e Simpson (1972). Esse domínio está associado aos processos do “fazer” e envolve os órgãos do sentido, perpassando pela percepção, movimentos corporais e comunicação não verbal. A escolha foi fundamentada devido a instituição já possuir processos bem estabelecidos para avaliação da cognição (provas online com resolução de itens de múltipla escolha relacionados à área de formação) e pela escassez de referências de outros estudos nessa linha, já que este prisma é tão relevante para verificação da competência profissional. Para que este domínio pudesse ser representado adequadamente, fez-se o uso de uma matriz de referência, apontando quais as capacidades do itinerário formativo seriam priorizadas na avaliação.

Para a realização de exames dessa magnitude, é necessário fundamentar-se em alguma teoria com respaldo na comunidade científica e que apresente resultados possíveis de comparabilidade entre os sujeitos. A Teoria de Resposta ao Item (TRI) apresenta fatores importantes que colaboram para sua elegibilidade, como a experiência brasileira em sistemas públicos de avaliação e comparação dos resultados entre estudantes. Andrade, Tavares e Valle (2000, p. 8) afirmam que a TRI é um conjunto de modelos matemáticos que visa representar a probabilidade que um indivíduo possui em acertar um determinado item (questão) em função da sua habilidade (traço latente) do respondente, sendo que quanto maior a habilidade, maior a probabilidade de acerto. Sendo assim, optou-se por essa abordagem metodológica para auxiliar na interpretação dos resultados dos desempenhos dos estudantes.

1.1 OBJETIVOS

Os objetivos dessa dissertação foram segmentados em objetivo geral e específicos, conforme detalhamento nas seções correspondentes.

1.1.1 Objetivo geral

Propor uma escala que permita avaliar o domínio psicomotor da aprendizagem, no contexto da educação profissional, à luz da Teoria de Resposta ao Item.

1.1.2 Objetivos específicos

São cinco os objetivos específicos:

- Investigar experiências empregadas na avaliação do domínio psicomotor utilizando a Teoria de Resposta ao Item como método de análise e interpretação de resultados.
- Caracterizar o experimento objeto de estudo quanto à elaboração da matriz de referência do constructo, instrumento de avaliação utilizado e processo de aplicação das provas práticas.
- Realizar a estimação dos parâmetros dos itens de avaliação e proficiência dos estudantes na perspectiva do domínio psicomotor da aprendizagem.
- Elaborar escala de níveis de desempenho de estudantes da educação profissional, que permita a identificação da qualidade da proficiência e interpretação pedagógica do domínio psicomotor da aprendizagem.
- Utilizar a escala construída na amostra selecionada para diagnóstico dos níveis de desempenho dos estudantes da educação profissional.

1.2 DEFINIÇÃO DO PROBLEMA

Esta dissertação apresenta como problema de pesquisa o seguinte questionamento: como elaborar uma escala que permita avaliar o domínio psicomotor da aprendizagem de estudantes da educação profissional?

1.3 JUSTIFICATIVA DO PROBLEMA

Este problema de pesquisa possui como principais justificativas a atualidade do tema, frente às recentes legislações e expansão da oferta de vagas nessa modalidade de ensino; o ineditismo no formato da avaliação e utilização em um sistema em larga escala, no qual um instrumento de avaliação dessa natureza ainda não é utilizado como procedimento em

nenhuma modalidade de ensino do país; a relevância do tema frente à importância social e política, uma vez que o governo precisa garantir a oferta de ensino de qualidade conforme previsto na legislação; e o interesse do autor, como fonte de aprimoramento profissional e aplicação dos resultados na instituição em que atua.

1.4 DELIMITAÇÃO DO TEMA

A proposta deste projeto está na aplicação da TRI em uma frente de estudo, que está correlacionada com a linha de pesquisa “métodos quantitativos em avaliação educacional” do Programa de Pós-graduação em Métodos e Gestão em Avaliação (PPGMGA) da Universidade Federal de Santa Catarina (UFSC), cujo foco principal é a avaliação educacional em larga escala. Mais especificamente, pretende-se pesquisar a aplicação da TRI no contexto da educação profissional, como ferramenta para aferição de proficiência dos estudantes em relação ao domínio psicomotor da aprendizagem. Devido à ausência de avaliações nessa proporção e a carência de parâmetros para essa modalidade de ensino no país, a proposta busca contribuir para a reflexão da temática e apontar uma metodologia viável de mensuração deste domínio, considerando suas peculiaridades, que poderá servir como subsídios para estimular o desenvolvimento de uma política pública de avaliação da educação profissional no Brasil.

Nesse intuito, limita-se o campo de pesquisa deste projeto no contexto da avaliação da educação profissional no Brasil, considerando principalmente os critérios legais previstos na Constituição Federal de 1988 e Lei 9.394/1996. O objeto da pesquisa direciona a aplicação da TRI em avaliações práticas de proficiência (focadas no “saber fazer” e não no “saber”), como uma nova abordagem e alternativa para os tradicionais testes cognitivos utilizados nas avaliações em larga escala da educação básica e superior. Trata-se aqui uma visão delimitada para o traço latente em questão, sem desprezar os efeitos dos outros dois domínios (cognitivo e afetivo), mas por estratégia de segmentação e visibilidade, optou-se por aprofundar sua discussão.

A pesquisa teve como base temporal a aplicação de instrumentos de coletas em tempo presente, uma vez que não há um histórico de análise sobre esse objeto. O campo de observação abrangerá como amostra estudantes da educação profissional (de nível médio), vinculados ao SENAI, pertencentes a todas as regiões geográfica do país, limitando-se em abordar apenas o Curso Técnico em Mecânica, tendo em vista o nível de detalhamento que se pretende chegar.

1.5 ESTRUTURA DA DISSERTAÇÃO

Além do primeiro capítulo introdutório, esta dissertação apresenta outros três. No segundo capítulo estrutura-se a fundamentação teórica que auxiliou na formulação do objetivo do trabalho, bem como subsidiou na definição da metodologia a ser empregada e na análise dos resultados. É subdividida por três subseções, sendo a primeira um retrospecto sobre a educação profissional do Brasil, abordando o contexto em que esta modalidade está inserida. A segunda busca investigar sobre o constructo alvo – domínio psicomotor da aprendizagem – trazendo as principais teorias que sustentam seus contornos. Por fim, a terceira aborda os aspectos mais relevantes sobre a TRI e seus impactos em avaliações dessa natureza, além da identificação de experiências anteriores associadas ao tema.

O terceiro capítulo versa sobre os procedimentos metodológicos envolvidos nesse trabalho, sobretudo a estratégia de investigação para alcance dos objetivos propostos. Para isso, estabelece-se a caracterização da pesquisa e apresenta-se uma proposta de *design*, pontuando todas as fases e suas particularidades. Também são retratadas a população e amostra, bem como o caminho percorrido para construção e aplicação dos instrumentos de avaliação. Também são descritos os métodos e autores que foram selecionados para o tratamento dos resultados e que contribuem para a construção da escala de desempenho e sua interpretação pedagógica.

O quarto e último capítulo destaca toda a análise e discussão dos resultados levantados pela pesquisa, envolvendo a descrição da amostra estudada, calibração dos parâmetros das evidências, estimação das proficiências dos estudantes, análise do instrumento quanto a validade e fidedignidade, a construção de uma escala de níveis de desempenho, sua interpretação pedagógica e aplicação real no grupo avaliado. Importante destacar que cada uma dessas etapas teve sua relevância no processo para construção do raciocínio que permitiu a entrega do objetivo geral. Seguem-se, ao término, as considerações finais e as sugestões para novos trabalhos.

2 FUNDAMENTAÇÃO TEÓRICA

Segundo Köche (2003) a fundamentação teórica é realizada por meio de consulta em fontes primárias e em referências secundárias, cujas fontes mais comuns são obras publicadas, livros, monografias, periódicos especializados e documentos/registros disponibilizados por institutos de pesquisa. Seu escopo está em acumular e sistematizar as ideias de maior destaque já produzidas nas ciências, elencando-as de modo organizado para posterior utilização do pesquisador. Vergara (1997) contribui sobre suas principais funções, dentre elas: permitir maior clareza sobre o tema pelo autor; auxiliar na formulação de hipóteses e suposições; indicar quais os métodos mais adequados para resolução do problema; e identificar caminhos para coleta e tratamento dos dados.

Desta forma, estruturou-se este referencial teórico em três seções, considerando numa primeira abordagem a contextualização da educação profissional no Brasil, pois os participantes do escopo deste estudo são estudantes oriundos dessa modalidade. Na sequência fez-se um levantamento do estado da arte sobre a avaliação do domínio psicomotor da aprendizagem, tema central nesta discussão, perpassando desde as contribuições teóricas dos autores sobre as taxonomias, instrumentos e estratégias, até as experiências mapeadas em nível internacional sobre aplicação deste tipo de teste em circunstâncias correlatas. Por fim, enumerou-se algumas contribuições sobre a TRI e seus principais aspectos, abordando temas como modelos matemáticos, pressupostos epistemológicos e construção de escalas de traços latentes. A seguir são aprofundadas cada uma destas temáticas.

2.1 EDUCAÇÃO PROFISSIONAL NO BRASIL

A educação brasileira, historicamente, sempre esteve ancorada em uma dualidade característica, atribuindo escolas diferentes para castas sociais distintas. De um lado, os hegemônicos da nobreza tinham à disposição uma educação básica propedêutica e com o intuito de preparação para acesso em cursos de nível superior. Do outro, os menos favorecidos economicamente e socialmente dispunham apenas da educação profissional, orientada exclusivamente ao ingresso no mercado de trabalho (KUENZER, 2007).

Para compreender melhor esse contexto é necessário um olhar reflexivo sobre a perspectiva histórica da educação profissional, sua evolução ao longo dos séculos, desde os períodos mais remotos da história até os dias atuais. Paralelo a isto, acompanhar o desenvolvimento do

arcabouço legal que sustenta sua prática, seu papel e suas diretrizes no território nacional até os dias atuais, conforme descrito na seção a seguir.

2.1.1 Histórico da educação profissional no Brasil

No início da civilização os homens transferiam seus conhecimentos de modo rudimentar e simples. Conforme descrito por Manfredi (2002), este processo baseava-se apenas na observação, prática e repetição de tarefas, onde eram ensinadas técnicas de produção de utensílios, aperfeiçoamento de ferramentas de pesca e caça, instrumentos de defesa e outros artigos que lhes fossem úteis e auxiliassem no cotidiano. Ele destaca ainda que nas populações pré-históricas e civilizações sucessoras, os saberes eram transferidos de geração em geração, baseados na lógica do erro/acerto e na imitação de saberes, acumulados pelas suas experiências históricas-culturais.

O autor completa afirmando que a associação tardia entre educação e trabalho é fruto das relações sociais presentes durante as Idades Antiga e Medieval, uma vez que as noções de trabalho “[...] vão se construindo e reconstruindo ao longo da história das sociedades humanas, variando de acordo com os modos de organização da população e de distribuição de riqueza e poder” (MANFREDI, 2002, p. 34). Aqui inicia o dualismo, demarcado pela centralização do poder nos senhores feudais e na igreja, enquanto os escravos e serventes eram marginalizados. O conhecimento, portanto, cabia somente aos “escolhidos”, sendo exclusivo das classes dominantes.

Com a chegada da Primeira Revolução Industrial (entre os séculos XVII e XVIII), o sistema feudal dá lugar ao capitalismo caracterizado pelo poder de compra da força de trabalho e dos recursos necessários à produção, ascendendo uma nova classe social: a burguesia. A máquina a vapor foi a propulsora dessa mudança, pois possibilitou o início do sistema fabril e demandou a introdução de funções intelectuais nos processos produtivos, que até então eram braçais. Impulsionados nessa direção, os países mais desenvolvidos iniciaram seus organismos nacionais de ensino frente as emergentes mudanças que a sociedade enfrentava. Todavia, a dicotomia já existente apenas se reconfigurou, pois enquanto as escolas profissionais que surgiam eram designadas aos que dispõem somente de sua força de trabalho – trabalhadores, as escolas de ciências e humanidades eram privativas dos detentores do capital – futuros dirigentes (SAVIANI, 2007).

O modelo extrativista – na contramão da industrialização acelerada do hemisfério norte – imperava no Brasil, não havendo interesse da Coroa

Portuguesa na implantação de fábricas (SANTOS, In: LOPES et al., 2000). Isso posto, Fonseca (1961) esclarece que durante boa parte do período de colonização houve a preparação dos índios e escravos como os primeiros “aprendizes de ofícios”. Antagônico a esse processo, a educação propedêutica era apenas para a elite, sendo que até o século XIX não existia proposta sistemática de ensino no país. Somente com a chegada da família real portuguesa, em 1808, é criado o “Colégio das Fábricas” por D. João VI, cujo objetivo era a educação de artistas e aprendizes oriundos de Portugal, fato este considerado como o início da educação profissional no Brasil (GARCIA, 2000).

Cabe destacar a função que o sistema jesuítico de educação desempenhou durante o Brasil Colônia, quando os padres além de catequizar os índios foram os primeiros a pensarem em sistematizar o ensino de práticas informais de qualificação para o trabalho:

Nas oficinas existentes nos colégios espalhados pelos diferentes pontos do Brasil, os irmãos-oficiais exerciam e ensinavam ofícios ligados a atividade de carpintaria, de ferraria, de construção de edifícios, embarcações, de pintura, de produção de tijolos, telhas, louças, etc., de fabricação de medicamentos, de fiação e de tecelagem (MANFREDI, 2002, p. 69).

Com a proclamação da independência em 1822, Saviani (2007) explica o escalonamento da produção manufatureira e o surgimento de sociedades civis denominadas de Liceus, vocacionadas para aprendizagem das artes e dos ofícios com conhecimentos relativos à agricultura e ao comércio. Importante destacar que mesmo com a chegada dessas estruturas, permanecia discriminada a mão-de-obra escrava praticada durante o período colonial. Nessa perspectiva, Ghiraldelli Jr. (2008) também colabora informando que entre os anos de 1840 e 1856 surgiram as “casas de educandos artífices”, inspirado na doutrina militar e foco de atendimento para órfãos e os desvalidos. O autor continua discorrendo sobre os eventos sucessores, como a abolição da escravatura e a Proclamação da República (1889) no Brasil, que permitiu a expansão do trabalho assalariado e desenvolvimento dos centros urbanos.

A primeira iniciativa legal de que se tem registro para a educação profissional é datada em 11 de setembro de 1906, quando o então Presidente do Estado do Rio de Janeiro (nomenclatura dos governadores da época) Nilo Peçanha promulgou o Decreto nº 787. Criava-se, portanto,

quatro escolas profissionais para o ensino de ofícios e aprendizagem agrícola. Anos seguintes, em 1909, o Estado assume essa responsabilidade e instituem-se 19 escolas de artes e ofícios, em diferentes unidades da federação, consideradas estas as precursoras da rede federal e estadual de ensino técnico. Kuenzer (2007, p. 27) sobressalta que esses locais possuíam “a finalidade moral de repressão: educar pelo trabalho, os órfãos, pobres, e desvalidos da sorte, retirando-os das ruas”. Porém, a escassez de professores qualificados, infraestrutura precária e metodologia respaldada apenas em empirismo influenciaram no fraco desempenho dessas instituições (SANTOS, In: LOPES et al., 2000).

Escott e Moraes (2012) acrescenta que a educação profissional insurge na perspectiva assistencialista, pois seu foco era amparar os pobres e órfãos carentes de condições sociais e econômicas. Apesar disso e da demanda emergente de qualificação de mão-de-obra, havia o receio que os donos dos estabelecimentos comerciais e industriais tinham em relação à escolarização operária, uma vez que “[...] era supérfluo e até perigoso ensinar a ler, escrever e, especialmente, fazer contas aos operários [...]” (MANACORDA, 1995, p. 287). Isto porque, na visão deles, a instrução poderia impulsionar uma revolução em luta de condições mais dignas de trabalho, sendo necessário apenas desenvolver noções técnicas, domínio do ofício e disposição para laborar.

Avançando no tempo, em 1927 teve-se a primeira tentativa para a formulação de uma Lei onde tornava-se obrigatório o ensino profissional no país, denominado “Projeto Fidélis Reis”. Três anos depois, em 1930, criou-se a primeira estrutura da administração pública com essa finalidade, a Inspetoria do Ensino Profissional Técnico, cujo propósito era supervisionar as Escolas de Aprendizes Artífices, vinculadas anteriormente ao Ministério da Agricultura (CUNHA, 2000). As décadas seguintes, 1930 e 1940, consolidaram a industrialização brasileira, resultando na promulgação da Constituição de 1937 que abordava:

O ensino pré-vocacional e profissional destinado às classes menos favorecidas é, em matéria de educação, o primeiro dever do Estado. Cumpre-lhe dar execução a esse dever, fundando institutos de ensino profissional e subsidiando os de iniciativa dos Estados, dos Municípios e dos indivíduos ou associações particulares e profissionais. É dever das indústrias e dos sindicatos econômicos criar, na esfera de sua especialidade, escolas de aprendizes, destinadas aos filhos de seus operários ou de seus

associados. A lei regulará o cumprimento desse dever e os poderes que caberão ao Estado sobre essas escolas, bem como os auxílios, facilidades e subsídios a lhes serem concedidos pelo poder público (BRASIL, 1937, art. 129).

No ano de 1942 foi a vez da Reforma Capanema com a coroação do bipartidarismo educacional brasileiro, oficializando o caminho elitizado: primário ao ginásio, ginásio ao colégio, depois qualquer curso superior. Já as classes populares percorriam apenas do primário aos cursos profissionalizantes disponíveis, concedendo acesso ao curso superior na mesma área do anterior. Entretanto, a reforma não cumpriu seu objetivo, pois a classe média não possuía interesse na educação profissional, optando apenas pelo secundário (propedêutico), enquanto os mais pobres não conseguiam frequentar sequer esse nível escolar (GHIRALDELLI JR., 2008).

Neste mesmo ano, por meio do Decreto-Lei nº 4.048, estabeleceu-se o SENAI para atender as demandas dos industriários, quando “os alunos passaram a receber um salário para estudar e iniciaram treinamento nas próprias empresas, o que tornou essa rede de ensino mais atrativa para as classes populares” (GHIRALDELLI JR., 2008, p. 88). Na continuação, em 1946 e pelo Decreto-Lei nº 8.621, instituiu-se o Serviço Nacional de Aprendizagem Comercial (SENAC) para atender os interesses dos empresários.

Somente em 1961 é publicada a primeira Lei de Diretrizes e Bases da Educação Brasileira (LDB), Lei nº 4.024/1961, porém não estava superada a dualidade estrutural das distintas modalidades de ensino, visto que havia projetos pedagógicos distintos para formação de trabalhadores instrumentais e intelectuais. Nessa lógica, Kuenzer (2007) prossegue afirmando que o legislativo passou a reconhecer saberes além do tradicional e integrou a educação profissional ao sistema regular de ensino, estabelecendo equivalência entre os cursos propedêuticos e laborais para prosseguimento dos estudos.

Continuando, em plena Revolução Civil Militar (iniciada em 1964), a Lei nº 5.692/71 tentou implantar a reforma educacional de nível médio profissionalizante para todos, momento este em que a habilitação profissional passou a ser prioridade em relação ao ensino propedêutico. Nesse interim, o governo argumentava sobre a necessidade incipiente de elevar a escolarização para um projeto de desenvolvimento econômico por meio de uma nova fase de industrialização conhecida como “Milagre Brasileiro”. A proposta era a “garantia” de inserção dos profissionais no

mercado de trabalho, porém a população de baixa renda não conseguia chegar ao segundo grau, promovendo a sua exclusão social (ESCOTT e MORAES, 2012).

Progredimos agora para a Constituição Federal de 1988, também conhecida como a “constituição cidadã”. Em seu terceiro capítulo, Seção I, estabeleceu a competência da União para legislar sobre as diretrizes e bases da educação nacional. No art. 214 faz menção à formação para o trabalho:

A lei estabelecerá o plano nacional de educação, de duração plurianual, visando à articulação e ao desenvolvimento do ensino em seus diversos níveis e à integração das ações do Poder Público que conduzam a: I - erradicação do analfabetismo; II - universalização do atendimento escolar; III - melhoria da qualidade de ensino; IV - formação para o trabalho; e V - promoção humanística, científica e tecnológica do País (BRASIL, 1988, art. 214).

Alguns anos se passaram e entrou em vigor a nova LDB instituída pela Lei nº 9.394/1996 que passou organizar a educação nos níveis escolares intitulados de Educação Básica e Educação Superior, sendo a primeira composta pela educação infantil, ensino fundamental e médio e a segunda pelos cursos de graduação e pós-graduação (tanto *Lato Sensu* quanto *Stricto Sensu*). Logo na sequência, o Decreto nº 2.208 de 17 de abril de 1997 regulamentou a educação profissional, integrando-a nas diferentes formas de educação e trabalho, desde os níveis mais elementares até o nível superior. Sua estrutura, portanto, abrangia os seguintes níveis: a) básico: voltado à qualificação e requalificação de trabalhadores independente de estudos anteriores; b) técnico: proposto à habilitação profissional para alunos egressos do ensino médio; e c) tecnológico: modalidade de nível superior na área tecnológica, direcionada aos alunos procedentes do ensino médio técnico (BRASIL, 1996; BRASIL, 1997; CANALI, 2009).

Canali (2009) alega que nessa mesma época foi implementado pelo Governo Federal o Programa de Expansão da Educação Profissional (PROEP). A estratégia estava diretamente ligada ao processo de privatização do estado brasileiro em consonância a política neoliberal. Os objetivos passavam pela criação de novas escolas de educação profissional no âmbito dos estados e municípios ou por meio de convênios

com a iniciativa privada, limitando a atuação da União apenas à rede federal (Centros de Formação de Educação Profissionais – CEFET).

O último marco que se registra aqui é a Lei nº 12.513/2011 que organizou o Pronatec, possibilitando a ampliação da oferta de cursos de educação profissional. Conforme redação, o foco era o atendimento de estudantes do ensino médio da rede pública e trabalhadores, e contemplava, dentre muitas, as seguintes ações: ampliação de vagas e expansão da rede federal de educação profissional e tecnológica; incentivo a ampliação de vagas e expansão da rede física dos serviços nacionais de aprendizagem; oferta de bolsas-formação; financiamento da educação profissional e tecnológica; articulação com o Sistema Nacional de Emprego (SNE).

Embora atualmente o dualismo estrutural da educação passou por mudanças e não está mais perceptível na legislação, o legado histórico ainda impacta essa modalidade. As políticas públicas são diminutas em comparação com a educação básica, por exemplo, carecendo de ações que promovam a sua valorização e regulação, ainda que haja uma diretriz no atual PNE. Para se ter real clareza da dimensão, os dados do “Movimento Todos pela Educação” apontam que, em 2017, foram realizadas 1.791.806 matrículas em educação profissional técnica, muito aquém das 5.224.584 necessárias para o atingimento desta meta (até 2024).

2.1.2 A educação profissional de nível técnico

A educação profissional é, portanto, uma modalidade complementar prevista na LDB de 1996 e seu principal objetivo é a formação para acesso ao mercado de trabalho, tanto de estudantes, quanto de profissionais que necessitam ampliar suas qualificações. Por ser uma modalidade transversal, está integrada em ambos os níveis de educação (Básica e Superior), podendo ser ofertada por cursos de: Formação Inicial e Continuada (FIC), Técnico de Nível Médio; ou Graduações Tecnológicas e Pós-graduações.

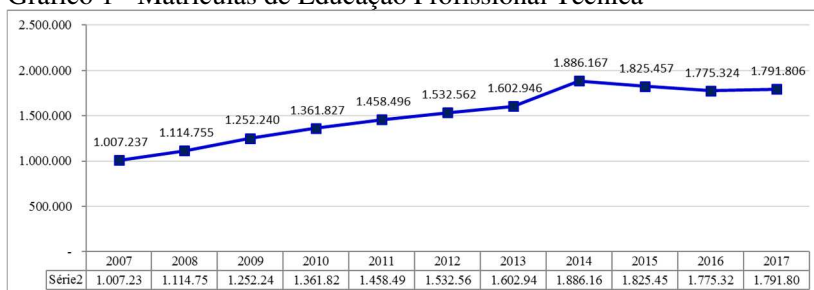
Especificamente sobre a habilitação técnica de nível médio, Freitas (2010) esclarece que são destinadas aos estudantes do ensino médio ou egressos, podendo ser realizada por instituições devidamente autorizadas pelas Secretarias Estaduais de Educação. Ainda é possível executar sua oferta de maneira integrada, concomitante ou sequencial ao ensino médio, conforme previsto na legislação.

A Resolução nº 6 do MEC, de 20 de setembro de 2012, detalha essa e demais regras, definindo as Diretrizes Curriculares Nacionais (DCN) para a educação profissional nesta modalidade. Em seu artigo 5º

deixa claro seu propósito está em “proporcionar ao estudante conhecimentos, saberes e competências profissionais necessários ao exercício profissional e da cidadania, com base nos fundamentos científico-tecnológicos, socio-históricos e culturais”. Também organiza os cursos por eixos tecnológicos elencando-os no Catálogo Nacional de Cursos Técnicos (CNCT), vinculando cada título a uma ou mais ocupações previstas na Classificação Brasileira de Ocupação (CBO). O mesmo instrumento jurídico também aborda sobre a avaliação da aprendizagem dos estudantes, que deverá ser contínua e cumulativa, com preferência dos aspectos qualitativos sobre os quantitativos, assim como dos resultados ao longo do processo sobre as eventuais provas finais.

Trazendo alguns números para contextualizar o que está sendo discutido, os dados do Observatório do PNE apontam a evolução de matrículas de educação profissional técnica dos últimos anos, conforme detalhado no Gráfico 1. Como é possível observar, as matrículas demonstram crescente expansão, apenas registrando queda a partir de 2015 devido à desmobilização do Pronatec e leve retomada em 2017 (último ano com dados disponíveis). Para que se tenha noção da proporção, os dados de 2017 revelaram que 47,7% das matrículas são na forma subsequente, enquanto 30,3% são integradas e apenas 17,9% concomitante (demais refere-se à modalidade FIC e Educação de Jovens e Adultos - EJA). Além disso, há certo equilíbrio entre as redes públicas e privadas, registrando a distribuição de 58,9% e 41,1%, respectivamente.

Gráfico 1 - Matrículas de Educação Profissional Técnica



Fonte: Observatório PNE (2019)

Após entender o histórico e cenário atual da educação profissional no país, busca-se compreender com mais profundidade o constructo a ser avaliado por essa pesquisa. Nesse sentido, aos estudantes vinculados a essa modalidade é relevante verificar o quanto consegue-se mobilizar na prática os conhecimentos adquiridos em sala de aula para aplicação em

situações reais e complexas no mundo do trabalho. Por esse motivo, o “saber fazer” é imprescindível nesse processo formativo, doravante aqui denominado de “domínio psicomotor da aprendizagem”, conforme exploração a seguir.

2.2 AVALIAÇÃO DO DOMÍNIO PSICOMOTOR DA APRENDIZAGEM

O termo avaliação vem do latim, *a+valere*, e tem por significado o valor ou mérito em relação a um objeto de pesquisa (KRAEMER, 2006). Luckesi (2011) caracteriza-a como um mecanismo de “ajuizamento da qualidade”, sendo, portanto, um julgamento do valor sobre uma determinada realidade para a tomada de decisão. Guba e Lincoln (1981) acrescentam, além do julgamento, a discriminação como característica elementar, pois visa o recolhimento de informações para atribuir um juízo de valor.

Dentre os múltiplos campos em que a avaliação é utilizada, a aprendizagem é um dos mais expressivos, podendo ser delimitada como um meio de adquirir dados sobre o progresso dos estudantes e retroalimentar o processo de ensino-aprendizagem. Orienta ainda o trabalho docente no planejamento das suas ações com vista em auxiliar o aluno na evolução da sua jornada formativa (FURLAN, 2007). Nessa concepção, busca-se identificar, medir, investigar e analisar o comportamento e o desempenho não só do estudante, mas do educador e da escola (SANT’ANNA, 1995).

Libâneo (1994) alerta sobre a dificuldade do ato de avaliar, até porque este processo não se resume apenas em aplicar provas e imputar notas. Se assim fosse, apenas possibilitaria indícios que necessitam de análise qualitativa pelo docente, sem que cumprisse o seu papel pedagógico-didático de diagnóstico e controle. Silva (2004) continua corroborando nesse sentido, afirmando que a interação entre ensino e aprendizagem passa pela intervenção consciente do professor, para que o estudante também possa compreender suas próprias potencialidades e fraquezas na sua trajetória escolar. Avaliar, conseqüentemente, incide numa observação constante para proceder com ações educativas que melhorem a condição do aluno e permita a reflexão do trabalho do docente quanto ao que foi proposto (HOFFMANN, 2008).

Para tanto, não é possível avaliar sem previamente planejar, conforme afirma Goldberg e Sousa (1979). É necessário estabelecer causa/efeito dos objetivos e promover o “controle de qualidade do planejamento”, funcionando como um termômetro e indicando possíveis

falhas no processo de ensinar e aprender. Ainda nessa temática, Santos *et al.* (2005) prosseguem alegando que os objetivos da aprendizagem devem ser avaliados continuamente e ao longo de todo o processo, voltados não só ao desempenho do aluno, mas também ao professor e o plano de ensino.

Em síntese, a avaliação educacional é uma ferramenta de fundamental importância no sistema escolar, utilizada para investigar os conhecimentos, atitudes ou capacidades que os alunos se apropriaram ou não. Revela não somente os objetivos de ensino (aqueles previstos nos currículos e itinerários formativos) já alcançados em determinado momento, mas, sobretudo, as dificuldades que devem ser trabalhadas para seu alcance (KRAEMER, 2006).

Quanto aos tipos de avaliação existentes nesse contexto, Santos *et al.* (2005) apresentam algumas possibilidades dentre as quais aqui se destacam três:

- diagnóstica – tem por finalidade a consulta daquilo que os estudantes aprenderam ou não previamente à exposição de um determinado conhecimento, permitindo que o docente mapeie as necessidades de aprendizagem;
- formativa – possui como objetivo a verificação do que foi ministrado aos alunos e tem como foco acompanhar o processo de ensino-aprendizagem quanto aos objetivos propostos no currículo; e
- somativa – seu propósito está em avaliar se o estudante está apto ou não a ser promovido de uma fase para outra, isto é, se no conjunto de tudo aquilo que foi ensinado o aluno conseguiu desenvolver as capacidades necessárias para progressão ou finalização dos estudos.

Independente da função escolhida, Morreto (2010) chama atenção para que a avaliação ocorra por meio de situações em que o estudante necessite mobilizar suas competências o mais próximo possível do que se solicitaria a um profissional (e não estudante). Nesse intuito, os próprios educandos teriam a oportunidade de rever seus pré-conceitos sobre o processo de avaliação, encarando-as como “momentos privilegiados de estudo e não como acerto de contas” (MORRETO, 2010, p.30). Para isso, precisa-se pensar numa lógica em avaliar por competências e não apenas conhecimentos isolados, conforme desdobramento a seguir.

2.2.1 Avaliação por competências

Nessa perspectiva, inicialmente explora-se o conceito de competência na literatura. Um dos principais expoentes nesse sentido é Perrenoud (2000), que a define como a faculdade de mobilizar recursos, sejam eles saberes, informações, habilidades e inteligências, para com efetividade enfrentar e resolver situações ou problemas em diferentes contextos. Le Boterf (2003) enfatiza que não se trata, portanto, de um estado nem da soma dos conhecimentos ou capacidades aprendidas pelo aluno, mas da sua aplicação numa situação específica, acionando os recursos e atentando-se as restrições de cada circunstância.

Para adentrar na discussão, Perrenoud (1999) destaca que nas formações por competência a metodologia de ensino-aprendizagem está ancorado em três pilares: conhecimentos, habilidades e atitudes, conforme detalhado na Figura 1. Nessa mesma linha, conforme sintetiza Carbone (2006), a competência é o conjunto desses três recursos indispensáveis para exercer uma determinada atividade, assim como expressa também o desempenho do indivíduo em um dado contexto.

Figura 1 - Componentes da competência

CONHECIMENTO	HABILIDADE	ATITUDE
<ul style="list-style-type: none"> • Refere-se ao conjunto de informações necessários para realizar uma atividade. • Está relacionada ao "saber". 	<ul style="list-style-type: none"> • Refere-se as ações e funções necessárias para realizar uma atividade. • Está relacionada ao "saber fazer". 	<ul style="list-style-type: none"> • Refere-se a postura em que o indivíduo deverá demonstrar ao realizar uma atividade. • Está relacionada ao "saber ser".

Fonte: elaborado pelo autor (adaptado de Perrenoud, 1999)

Transferindo esse conceito para o mundo do trabalho, tem-se a competência profissional como a mobilização desses mesmos componentes para o desempenho esperado (necessidades da organização) em relação ao trabalhador, quanto às demandas dos cargos ou funções existentes na estrutura organizacional. Em outras palavras, a competência é percebida como um “estoque de recursos” que, quando combinados, potencializam o modo como os indivíduos realizam as diversas práticas profissionais. Na educação profissional, esses recursos estão associados aos domínios cognitivo, psicomotor e afetivo da aprendizagem (FLEURY E FLEURY, 2001, p. 185; FANEMA, 2004; DEPRESBITERIS, 2005).

Depresbiteris (2005) aborda ainda que nesse mesmo contexto o termo competência surgiu inicialmente em países desenvolvidos na

tentativa de conectar os sistemas educativos com os produtivos. As mudanças trazidas pelos avanços científicos e tecnológicos promoveram a necessidade de um novo perfil de trabalhador, com maior autonomia e iniciativa, além da capacidade de resolução de problemas. Por isso, na visão dela, a formação deve ser focada em mecanismos que possibilitem o educando a pensar e agir, preparando-o para um universo de mudanças aceleradas.

Logo, a competência não é algo que venha do berço. Todos os indivíduos possuem potencialidades para desenvolvê-las, transpassando pela aprendizagem significativa e priorizando a aplicação dos conteúdos formativos relevantes e que façam sentido no contexto do trabalho. Assim, o ensino por competências deve primeiramente desafiar o aluno por meio de uma situação complexa para que, somente em um segundo momento, estabelecer os conhecimentos necessários para sua abordagem.

Consequentemente, a avaliação da competência passa necessariamente pela observação do desempenho dos alunos nas atividades executadas, aplicando critérios que correspondem ao juízo de valor da adequação ou não aos objetivos propostos (no projeto do curso, por exemplo). Os instrumentos avaliativos, em especial na educação profissional, devem considerar sobretudo as habilidades psicomotoras, demandando que o aluno “mostre como faz” determinada tarefa (FEUERWERKER, 2002).

Algumas preocupações então surgem em relação aos caminhos em que avaliação por competências possa percorrer, como alerta Depresbiteris (2005) sobre a dicotomia: ser ou não ser competente como diagnóstico final. Não se pode pensar apenas no caráter somativo, mas valorizar principalmente o formativo, isto é, entender que o processo de aprendizagem é fluído, passível de correção e redirecionamento. Compreender que nessa vertente a competência é mutável, dinâmica e que necessita de constante revisão frente às exigências laborais. Todavia, a autora ajuíza “podemos avaliar os recursos que serão mobilizados no desenvolvimento da competência, mas essa competência só será avaliada na realidade do trabalho” (DEPRESBITERIS, 2005, p. 10).

Para adentrar um pouco mais sobre as “habilidades psicomotoras” conforme citado por Feuerwerker (2002), há de se explorar sobre os três domínios da aprendizagem, conforme fundamentação teórica levantada na próxima seção.

2.2.2 Domínios da aprendizagem

Uma avaliação por competência deve levar em consideração a mobilização de recursos estabelecidos por conhecimentos, habilidades e atitudes, conforme visto anteriormente. Estas múltiplas dimensões, no contexto da aprendizagem, podem ser traduzidas em domínios, doravante nomeados de cognitivo, psicomotor e afetivo, que segundo Bloom (1983) e Ferraz e Belhot (2010) podem ser compreendidos como:

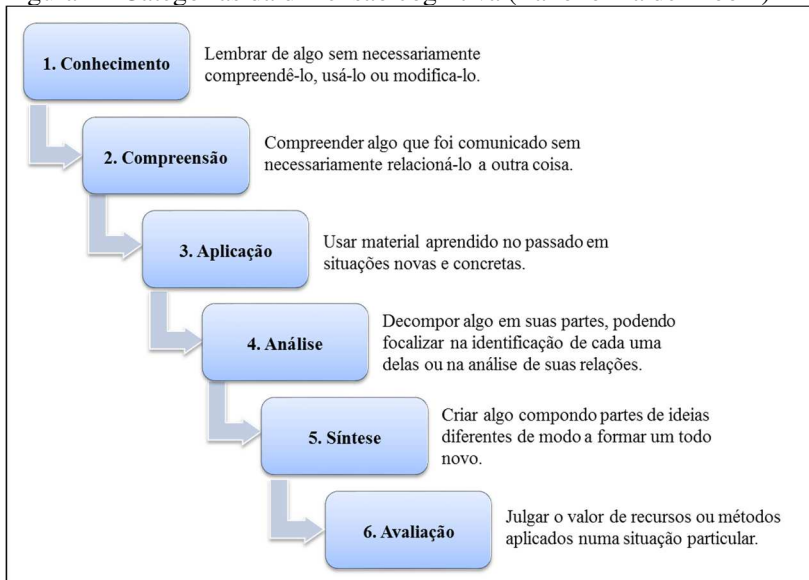
- cognitivo – relacionado ao modo como os indivíduos aprendem o “saber” de um determinado conteúdo, como os compreendem e analisam, isso é, o raciocínio e processo intelectual percorrido pelo estudante na aquisição dos novos conhecimentos;
- psicomotor – associado às habilidades do “fazer”, envolvendo os órgãos do sentido e ativação neuromuscular para a realização de atividades específicas, passando pela percepção, movimentos e comunicação não verbal; e
- afetivo – ligado à área emocional e, portanto, ao “ser”, representando as atitudes, crenças, valores e juízos sobre os sujeitos e objetos que orientam o estudante em comportamentos específicos durante a realização das tarefas.

A composição desses três domínios justapõe-se e foram inicialmente identificadas por um grupo de psicólogos liderados por Benjamin Bloom, com posterior publicação no trabalho “*Bloom’s Taxonomy of the Cognitive Domain*” (BLOOM, 1983). A “Taxonomia de Bloom”, como ficou conhecida, inicialmente pensada para definição de objetivos educacionais, é uma propícia ferramenta para planejar a avaliação do aprendizado, em especial nas metodologias de ensino construtivistas (BOOKER, 2007). Cabe destacar ainda que na prática os domínios não se manifestam separadamente durante a aprendizagem dos estudantes, sendo esta divisão apenas ilustrativa e puramente didática, com o propósito de instrumentalizar e orientar os processos educacionais (ABBAD *et al.*, 2006).

Especificamente tratando do domínio cognitivo, Bloom (1983) pôde perceber que mesmo possuindo condições similares de ensino (refutando variáveis externas) todos os estudantes aprendem, mas se diferenciam no nível de aprofundamento e fixação do conhecimento. O processo cognitivo é ativado de acordo com as associações mentais de

cada indivíduo, respeitando os diferentes tempos e percorrendo seis etapas hierarquicamente mapeadas, conforme demonstrado na Figura 2.

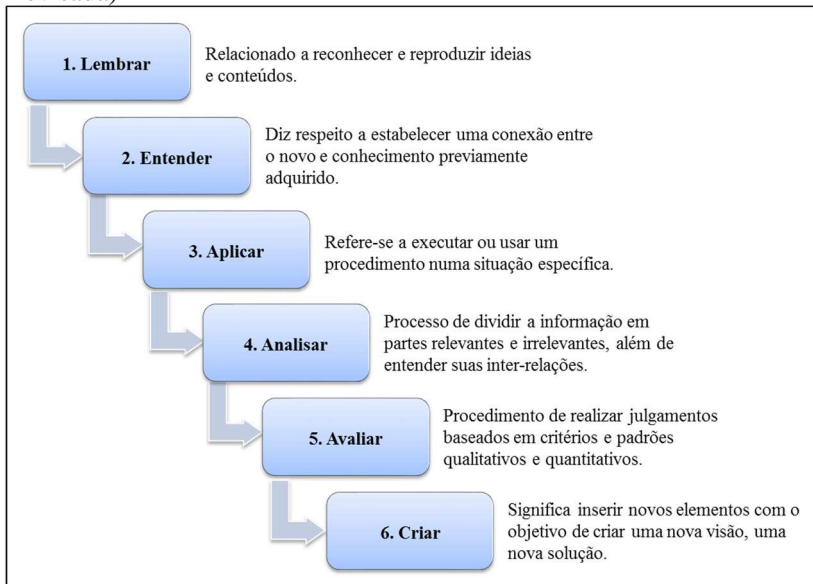
Figura 2 - Categorias da dimensão cognitiva (Taxonomia de Bloom)



Fonte: elaborado pelo autor (adaptado de Bloom, 1983)

Com base na aplicação da metodologia por diversos autores e os avanços pedagógicos/tecnológicos que ocorreram ao longo da história, após alguns anos da publicação houve uma reestruturação liderada por David Krathwohl (que participou junto com Bloom da primeira versão). A “Taxonomia de Bloom Revisada”, como passou a ser pronunciada, mantém as seis categorias, porém reclassificadas e com nomenclaturas distintas. As principais modificações realizadas forma: a categoria “conhecimento” permaneceu no primeiro nível, mas passou a ser denominada de “Lembrar”; “Compreensão” transformou-se em “Entender”; as três próximas categorias tiveram sua escrita modificada para o verbo no infinitivo, ficando “Aplicar”, “Analisar” e “Sintetizar”; e a última categoria de “Avaliação” converteu-se para “Criar”, além das duas últimas categorias serem invertidas, conforme se pode visualizar na Figura 3. Essas mudanças foram incorporadas porque as novas designações expressavam com mais exatidão o resultado esperado de um determinado estímulo (ANDERSON & KRATHWOHL et. al, 2001).

Figura 3 - Categorias da dimensão cognitiva (Taxonomia de Bloom Revisada)



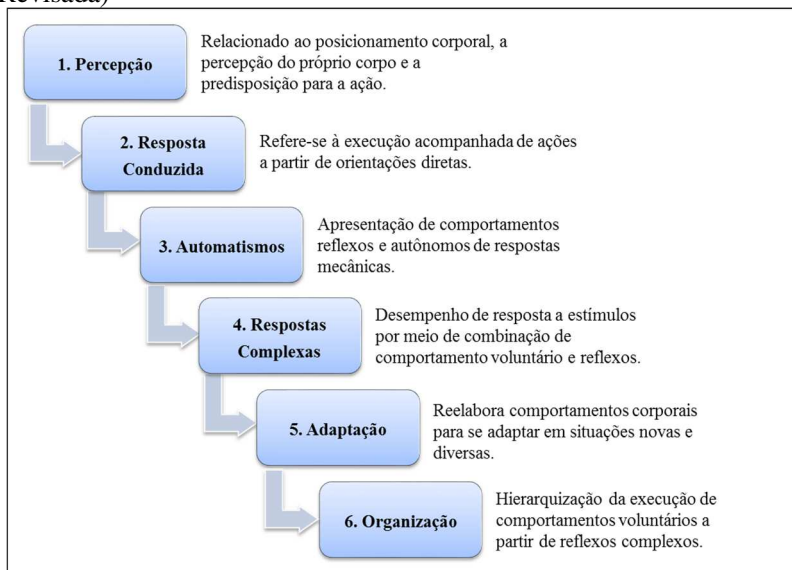
Fonte: elaborado pelo autor (adaptado de Anderson & Krathwoh et. al, 2001)

A Taxonomia de Bloom (seja ela a original ou revisada) é uma das principais contribuições para o meio acadêmico, pois desvela caminhos para que os docentes estimulem a aprendizagem dos estudantes, conscientizando-se dos níveis hierárquicos percorridos, sejam eles cognitivo, psicomotor ou afetivo (CONKLIN, 2005).

Bloom não descreveu inicialmente as categorizações para os domínios psicomotor e afetivo, porém outros autores desafiaram-se a identificar seus estágios. Quanto ao domínio psicomotor, que representa a mobilização das capacidades para a execução das tarefas e que demandem a motricidade e organismo muscular, apresenta-se também seis níveis, conforme detalhamento na Figura 4.

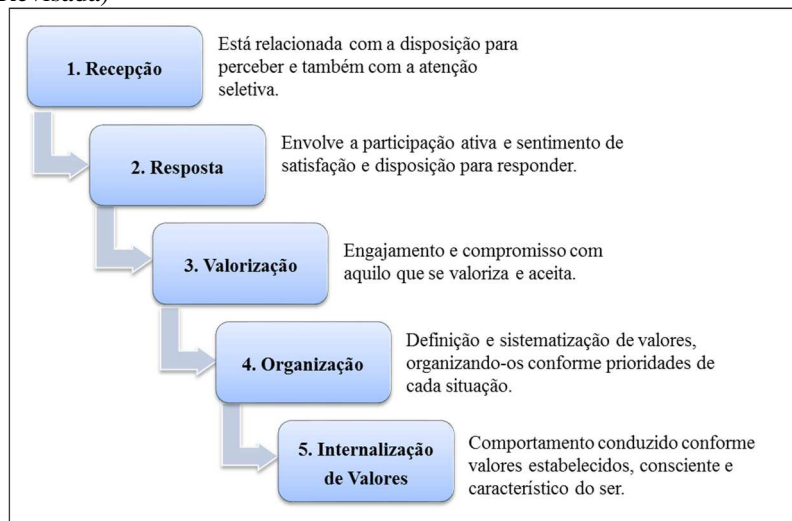
De mesmo modo, buscou-se definir uma sequência lógica de como se desenvolve o domínio afetivo, estruturando-se em cinco níveis como especificado na Figura 5.

Figura 4 - Categorias da dimensão psicomotora (Taxonomia de Bloom Revisada)



Fonte: elaborado pelo autor (adaptado de Dave, 1970 e Simpson, 1972)

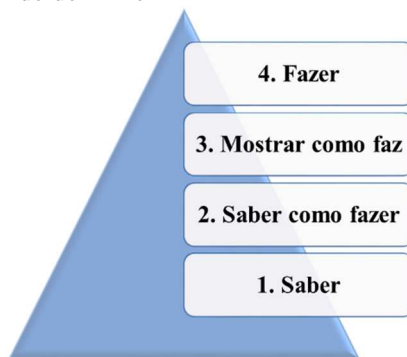
Figura 5 - Categorias da dimensão afetiva (Taxonomia de Bloom Revisada)



Fonte: elaborado pelo autor (adaptado Krathwohl, 2002)

Outra contribuição é advinda de George Miller, estudioso da educação médica e de métodos de avaliação aplicados ao ensino das profissões da saúde. Ele desenvolveu um modelo conceitual denominado de “Pirâmide de Miller”, elucidando as bases cognitivas e habilidades para verificação da prática profissional. Embora voltado às áreas clínicas, sua teoria vem ganhando bastante espaço no meio educacional, sendo representada graficamente conforme demonstrado na Figura 6 (MILLER, 1990).

Figura 6 - Pirâmide de Miller



Fonte: elaborado pelo autor (adaptado de Miller, 1990)

O autor classificou os dois primeiros extratos da pirâmide pertencentes ao domínio cognitivo, devendo estes serem avaliados conforme aferição da aquisição de conhecimentos. Todavia, a natureza do conhecimento é diferente, pois enquanto o primeiro nível está voltado ao aspecto teórico, isso é, se o estudante “sabe” os fatos, o segundo “saber como fazer” já é focado na aplicação. Em ambos os casos se sugere alguns instrumentos de avaliação, como os testes de múltipla escolha, questões dissertativas e provas orais, porém na segunda etapa deve-se realizar a “contextualização clínica”.

O estágio “mostrar como faz” corresponde ao domínio psicomotor e volta-se para a avaliação das habilidades e competências técnicas necessárias à profissão. O principal instrumento a ser utilizado aqui são as provas práticas que simulam tarefas do contexto do trabalho, sendo essencial a observação direta do avaliado pelo avaliador durante a execução do processo.

Já a última instância referente ao “fazer” e integra os três domínios da aprendizagem, verificando se a competência foi desenvolvida no próprio ambiente de trabalho. É quando o estudante já formado é avaliado

em relação ao desempenho da própria profissão, isto é, no momento em que está fazendo e exercendo sua prática profissional, porém ainda sob supervisão. Na área médica é comum esse procedimento, uma vez que o residente é testado quanto ao atendimento das pessoas assistidas e seus encaminhamentos clínicos reais.

Destaca-se que a contribuição de Miller é, sobretudo, apontar ao docente seus objetivos educacionais, porém sem restringir ao conhecimento fático, mas na aplicação prática do que se é estudado. Isso demanda uma mudança no processo formativo, necessitando inserir o educando em laboratórios e ambientes simulados para que ele possa praticar e aplicar legitimamente as capacidades requeridas pelo seu contexto de trabalho (COLES, 2002).

Harrow (1983) também apresentou taxonomia para o descrever o domínio psicomotor e também definiu seis categorias: (1) movimentos reflexivos; (2) movimentos fundamentais básicos; (3) movimentos de percepção; (4) atividades físicas; (5) movimentos hábeis; e (6) comunicação não-discursiva. No entanto, seu trabalho estava mais orientado para o desenvolvimento da motricidade no processo de crescimento da criança e suas interações psicomotoras nas atividades físicas.

Diante das características particulares de cada uma das propostas percorridas, optou-se por considerar a adaptação dos trabalhos de Dave (1970) e Simpson (1972) como espectro de análise em relação ao instrumento de avaliação adotado, pois foi o que mais se aproximou das demandas requisitadas pela educação profissional. Além disso, buscou-se também desenvolver uma reflexão a respeito de como a Pirâmide de Miller (1990) pode ser aplicada nesse contexto.

Portanto, o entendimento dos domínios da aprendizagem auxilia não apenas a definir os objetivos educacionais, mas também a avançar no entendimento do que precisa ser avaliado numa perspectiva por competência. Conforme Figari (1996) esse é o processo de “referencialização” e significa estabelecer os critérios que vão orientar o que os estudantes devem de fato ter desenvolvido. Uma alternativa para metodizar isso são as matrizes de referência, conforme abordado a seguir.

2.2.3 Matriz de referência para avaliações em larga escala

Araújo e Rabelo (2015) alegam que a sistematização de forma coerente e planejada destes referenciais ocorre por meio da utilização de um instrumento metodológico (matriz de referência) cujo principal objetivo é a identificação de competências. No campo educacional foca-

se no perfil da formação esperado e nos itinerários cognitivo, psicomotor e afetivo que o estudante deverá percorrer para desenvolvê-lo. Por esse motivo o nome “matriz” é atribuído, devido a diagramação do cruzamento das atividades profissionais com os recursos que deverão ser mobilizados para sua realização, configurando-se como as metas característica da formação.

Destarte, não se pode confundir matriz de referência com matriz de ensino/curricular. Enquanto a primeira busca avaliar os objetivos educacionais estabelecidos pelo perfil de conclusão, a segunda caracteriza-se como um guia para operacionalização da proposta pedagógica, ou seja, um norteador do trabalho docente para o processo formativo (NERY, 2000). Nesse sentido, este artefato é o principal norteador de um processo avaliativo, pois deixa claro o que pretende ser avaliado e serve como balizador para a construção dos itens (KLEIN, 2013).

Há de se registrar sobre a importância do processo avaliativo em larga escala como incentivo à gestão democrática e cumprimento dos aspectos legais. Cabe à União, conforme a LDB (BRASIL, 1996), artigo 4º, inciso IX “garantir padrões mínimos de qualidade de ensino [...]”. O objetivo, portanto, é a obtenção de dados e sua análise para diagnosticar a situação da educação e permitir a implantação, manutenção e reformulação de políticas educacionais.

Algumas características são apresentadas por Souza e Oliveira (2003) sobre este tipo em especial: normalmente são externas às redes de ensino ou à instituição a qual se destina; utiliza-se de testes padronizados aplicados a uma população de grande porte; é ancorada em estatísticas que garantem a credibilidade dos resultados, como a TRI; e permitem a comparabilidade dos dados por meio da sequência histórica, para levantamento e tratamento dos resultados.

Os autores esclarecem ainda sobre um risco iminente que pode ocorrer a este processo: a valorização de procedimentos competitivos. Muitos sistemas pontuam e ranqueiam seus desempenhos em escalas comparativas para punir ou premiar, redundando em critérios para alocação de recursos e remunerações diferenciadas, citado como exemplo. Fischer (2010) esclarece definitivamente que a finalidade fundamental desse processo é a tomada de decisões pedagógicas e administrativas com base nos referenciais estabelecidos (definidos na matriz de referência).

Uma próxima etapa é uma contemplação sobre as estratégias de escolha e utilização dos instrumentos de avaliação adequados as

características de um teste com base em competências, fato este que passa a ser discutido na próxima subseção.

2.2.4 Instrumentos de avaliação

Inicialmente precisa-se compreender a importância do instrumento avaliativo como mecanismo para aferição da aprendizagem dos estudantes. Segundo Abbad *et al.* (2006), a construção dessas medidas deve levar em consideração a modalidade, recursos disponíveis, perfil dos participantes, múltiplas fontes de avaliação, uso de variados tipos de itens e padronização dos instrumentos. Além disso, na perspectiva da educação por competência, como não há um único método que congregue todos os elementos (conhecimentos, habilidades e atitudes), somente a combinação de diferentes instrumentos possibilita os resultados de uma avaliação dessa magnitude (NORCINI, 2010). Luckesi (2011) ainda reforça sobre a necessidade de o docente articulá-lo com os objetivos educacionais propostos e que serão avaliados, harmonizando o que foi ensinado com o que foi aprendido.

Gronlund (1985) sugere que devam ser observados três atributos elementares no processo de construção: validade, fidedignidade e viabilidade, conforme detalha:

- Validade – está relacionada com o que efetivamente pretende-se avaliar, abrangendo os aspectos de conteúdo, estatística e do constructo. Como não é possível verificar um currículo na integralidade ou todos os tópicos de relevância, adota-se o uso de uma matriz de referência que seleciona a validade do que de fato será mensurado.
- Fidedignidade – está conectada à precisão e objetividade do instrumento para garantir a segurança e coerência dos resultados obtidos. Ser fidedigno representa o controle do maior número de variáveis, sejam elas relacionadas ao avaliador (subjetividade, grau de capacitação), ao avaliado (experiência no método, conhecimentos prévios) ou mesmo a metodologia (ambiente, materiais).
- Viabilidade – concernente à execução (aplicação do teste) e a aceitação por parte de todos os envolvidos. Carece de observar alguns aspectos como as condições de tempo, espaço e materiais necessários conforme o planejamento da avaliação e o controle dos custos.

Na perspectiva da educação por competências, a articulação entre teoria e prática é de fundamental importância para avaliar se o estudante realmente aprendeu. O termo “teoria” vem do grego e significa “a viagem de uma missão festiva aos lugares de sacrifício”, o que leva ao sentido de observar, estudar e refletir. Já a expressão “prática” é derivada das palavras *práxis* e *práxeos*, também do grego, e constitui em uma ação consciente, abrangendo a ação produtiva com a atividade profissional. Embora possam ser considerados polos separados (mas não opostos), sendo uma a aplicação da outra, na visão dicotômica cabe à primeira pensar, planejar e elaborar, enquanto a segunda se preocupa em executar, agir e fazer. Todavia a visão aqui é de indissociabilidade e não de identidade, onde o desenvolvimento da competência só é possível quando as duas ficam evidenciadas (GORGEN, 1979; CANDAU e LELIS, 2003, p. 60-65).

Embora muitos sistemas avaliativos se preocupem com o domínio cognitivo e, conseqüentemente, com os testes objetivos ou dissertativos, na educação profissional é essencial verificar se os domínios psicomotor e afetivo também estão desenvolvidos, porquanto somente é possível determinar a competência com a mobilização dos três. Nessa acepção, visando ampliar o escopo de observação, a avaliação prática combinada com os testes descritivos pode ser uma solução para esse desafio. Ainda que na literatura não haja uma definição consolidada, entende-se como prova prática o teste que expõe o estudante a uma situação-problema que demande a busca pela sua resolução no âmbito do “fazer”, isso é, da aplicação dos recursos cognitivos para a entrega de um produto ou serviço.

O conceito de situação-problema é chave nesse processo, podendo ser compreendido como uma situação didática em que se propõe ao estudante a realização de uma tarefa que somente será efetivada quando superados seus desafios (MEIRIEU, 1998). Especificamente sobre o termo problema, este deve ser entendido como “aquilo que se enfrenta e cuja solução, já conhecida ou incorporada, não é suficiente ao menos como conteúdo” (MACEDO, 2005, p. 15).

Perrenoud (2000) destaca ainda que a situação-problema é um recorte da realidade e seu principal objetivo é desafiar o estudante a transpor um obstáculo para realizar uma determinada aprendizagem. Esse desafio deve ser bem ajustado conforme as possibilidades dos estudantes, pois de acordo com o apresentado por Meirieu (1998) não deve ser fácil o suficiente para a resolução imediata, nem difícil o bastante para a completa desmotivação e abandono. Logo, os estudantes deverão compreender o contexto, interpretá-lo, anexar informações adicionais que

auxiliem na sua resolução, avaliar as possíveis opções de solução e posteriormente apresentar suas conclusões (DELISLE, 1997).

Mais que um momento avaliativo, a utilização da situação-problema permite que os estudantes possam também aprender, uma vez que são desafiados por questões que requerem análise, interpretação, tomada de decisão, ou seja, intrigantes. Por esse motivo deve-se olhar muito mais pelo obstáculo a ser superado do que pelo grau de dificuldade, até porque que o desafio é uma decisão do elaborador (avaliador), enquanto o grau de dificuldade é relativo ao nível de traço latente dos estudantes (MACEDO, 2002; MEIRIEU, 1998).

Para escrever uma situação-problema, Jonassen (1992) explica que o autor deverá buscar inspirações no mundo real, ou seja, em casos próximos aos que o profissional que se pretende avaliar teria que resolver se estivesse no contexto do trabalho. Perrenoud (2000) e Lima e Linhares (2008) contribuem apresentando algumas observações importantes no momento da elaboração do problema:

- organizar-se em torno da resolução de um obstáculo, previamente bem identificado e motivando-os para a sua resolução;
- apresentar caráter concreto e que permita a articulação de hipóteses e conjecturas por parte do estudante, o levantamento de questionamentos e novas ideias;
- desafiar o estudante para que este não disponha de todos os meios para alcançar a solução imediatamente, sendo a necessidade de resolução o que mobiliza-o na busca dos recursos;
- estar ao alcance dos estudantes, operando numa “zona próxima”, propícia ao desenvolvimento intelectual;
- ancorar-se nos objetivos educacionais a que se propõem, limitando-se ao tempo disponível para a resolução; e
- amparar-se dentro dos parâmetros curriculares, focando-se no tema do objeto e evitando assuntos gatilhos que funcionem como distratores.

Resolver problemas, portanto, requer o uso de estratégias, reflexão e tomada de decisão. Demanda raciocinar em cada etapa, desde a apreciação do problema, do seu caráter, da representação mental, organizando as informações disponíveis e alocando os recursos no tempo disponível, até a avaliação dos resultados alcançados (STERNBERG, 2000).

Uma ferramenta para aplicação de situações-problema é a simulação, a qual vem ganhando espaço na esfera educacional nos últimos 40 anos. São empregadas em circunstâncias que exijam o treino ou obtenção de capacidades psicomotoras, onde o foco é a habilidade técnica. Devem ser executadas em ambientes que imitem a realidade e solicitem a solução de tarefas significativas, dispondo de recursos reais como máquinas, equipamentos e ferramentas (PAZIN FILHO e SCARPELINI, 2007; VANCE e BOSWORTH, 2003).

As principais vantagens do uso de simulações no processo de ensino aprendizagem, segundo Vance e Bosworth (2003), são:

- permitem que a aprendizagem alcance níveis mais profundos, pois os estudantes podem simular diversas respostas para um mesmo problema;
- experienciam o erro sem que haja consequências de elevado valor ou de caráter irreversíveis, já que o ambiente é controlado e de baixo risco;
- possibilita o “mergulho” em profundidade por ser muito próximo da realidade, envolvendo emocionalmente os estudantes.

Evidentemente que essa técnica também pode ser um recurso a ser utilizado nas avaliações do aprendizado, principalmente quando estamos falando de formação profissional. Especialmente pelo seu caráter prático, esse método adere-se bem nas avaliações por competência e não só de conhecimento, como nos testes cognitivos (MILLER, 1990).

Mas há de se observar que os produtos desenvolvidos pelos discentes, seja na situação problema ou na simulação, não é o bastante para averiguar se a competência foi realmente desenvolvida, conforme advertido por Depresbiteris (2005). É imprescindível observar o processo de execução das atividades, o modo com o estudante realiza as tarefas. Na visão da autora a avaliação também tem que ser competente, no sentido de sempre buscar contextualizar as situações e os atores envolvidos, considerando a avaliação prática um instrumento adequado para essa finalidade.

Porém como determinar se os resultados alcançados pelos estudantes são satisfatórios? Quais critérios que embasam o julgamento do docente para estabelecer quando uma capacidade foi plenamente desenvolvida? Para mitigar esses questionamentos e promover uma mensuração sem viés subjetivo, a utilização de rubricas ou evidências (como retratado por essa pesquisa) minimizam impactos e proporcionam

maior clareza sobre esse processo, tema este que passa a ser aprofundado agora.

2.2.5 Avaliar por meio de evidências

Na abordagem da avaliação por competências, necessariamente, devem ser pautados critérios objetivos que permitam o julgamento do desempenho do estudante, sejam eles relacionados aos saberes (conhecimentos), saber fazer (habilidades) e saber ser (atitudes). Não deverão ser alicerçados em invenções ou subjetividades, sendo importante definir referenciais concretos ou explícitos de modo a tornar inequívocas as “regras do jogo”, ou seja, quando o aluno atingiu ou não o que se pretende avaliar (DEPRESBITERIS, 2001; ALLAL, 1986).

Esses parâmetros incidem sobre a verificação da aquisição de conteúdos frutos do pensamento e do domínio das capacidades. Nessa perspectiva surge a expressão “juízo de valor”, empregada em função das definições e padrões pré-estabelecidos com base em escalas apropriadas, em comparação a critérios qualitativos ou quantitativos (HADJI, 1994; POPHAN, 1975). Ferraz *et al.* (1994) conclui que estes artifícios se referem ao indivíduo, tarefa ou condições em que se realiza, entre o valor realizado e o esperado.

Os termos critérios, parâmetros e padrões são usados no contexto da avaliação como palavras de sentido semelhante a uma base de referência para julgamento, optando por nomear de “evidência” no contexto deste trabalho, representando o resultado alcançado pelo aluno durante a realização de uma tarefa avaliativa. Hadji (1994) ensina que “referência” vem do latim, e significa “reportar”, logo precisa-se embasar-se em algo pré-existente, fundamentando o juízo de valor.

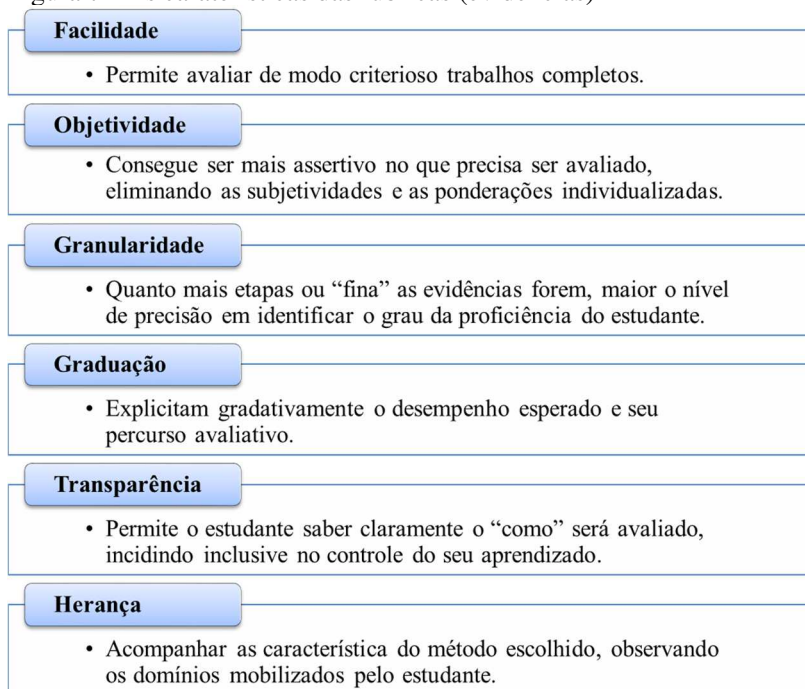
Quanto à natureza, poderão ser de caráter qualitativo ou quantitativo, conforme demonstrado por Depresbiteris (1989). Enquanto o primeiro diz respeito à descrição da qualidade do desempenho esperado, o segundo mensura por indicadores numéricos o quanto o estudante deverá alcançar para evidenciar que a capacidade foi desenvolvida. A autora ainda reforça que ambos devem ser atribuídos de modo complementar, pois apresentam objetivos diferentes.

Na literatura não é comum encontrar o termo “evidência” como unidade de medida, mas pode-se apoiar na nomenclatura de “rubricas”. Para Taggart *et al.* (2001), a palavra deriva do inglês – *rules* – que significa regras, representando os princípios estabelecidos e acordados entre o docente e os alunos para mensurar a sua aprendizagem. Busching (1998) define-a como uma ferramenta de orientação do trabalho do

professor durante o processo avaliativo, fornecendo parâmetros para identificar o nível de proficiência dos estudantes em cada tarefa. Para ele esse método aumenta a consistência da avaliação, até porque sugere de modo objetivo o que se necessita verificar.

Dinham e Roew (2009), por sua vez, apresentam algumas vantagens na utilização de rubricas: melhora o desempenho do estudante, porque aponta as expectativas e como satisfazê-las; torna o processo avaliativo mais objetivo e consistente; proporciona ao aluno uma prova mais informativa; promove que os docentes sejam mais assertivos quanto aos objetivos educacionais e qualidade do trabalho dos estudantes. Ainda nesse sentido, devem ser observadas algumas características para sua elaboração, conforme detalhamento na Figura 7.

Figura 7 - As características das rubricas (evidências)



Fonte: elaborado pelo autor (adaptado de Ludke, 2004)

O uso das rubricas ainda provoca outro aspecto bastante interessante, conforme também indicado por Busching (1998): a orientação do aluno no que tange ao esperado na avaliação. Isso porque

quando o estudante sabe claramente o que ele deve fazer não há presença de retrabalhos ou esquecimentos, além de poder conferir qual item contará maior pontuação, ou seja, quais os aspectos mais relevantes da prova.

Porém estabelecer esses critérios é uma tarefa árdua conforme sinaliza Turra *et al.* (1998), pois demanda clareza absoluta dos atributos indicativos do julgamento, de maneira estritamente objetiva e precisa, onde o resultado da avaliação é a síntese das evidências alcançadas. Em oposição, Luckesi (2011) diz que “dificilmente os professores definem com clareza [...] qual é o padrão de qualidade que se espera da conduta de um aluno, após ser submetido a uma determinada aprendizagem” (p.44). Encontrar esse equilíbrio resulta nos critérios mínimos a serem observados e faz com que as aprovações ou reprovações não estejam mais baseadas na arbitrariedade do professor, mas sim em fundamentos do comportamento da aprendizagem.

Após exposição sobre a mensuração do domínio psicomotor da aprendizagem, sobretudo na perspectiva da competência e do embasamento em evidências que comprovem sua real efetividade, buscou-se uma abordagem nas teorias de avaliação que amparasse estatisticamente a confiabilidade da medida e possibilitasse maior compreensão sobre o fenômeno. Nesse sentido, optou-se pela TRI como proposta devido às suas propriedades essenciais, as quais serão caracterizadas no próximo tópico.

2.3 TEORIA DE RESPOSTA AO ITEM

Para fundamentar esta designação, inicialmente é necessário entender um pouco da trajetória da psicometria e seu histórico junto as teorias de avaliação. No início do século XX, Spearman (1904) desenvolveu um conjunto de métodos e conceitos que seria nomeado posteriormente como a Teoria Clássica das Medidas (TCM) ou Teoria Clássica do Teste (TCT). Na sequência, Thurstone (1994) continuou a descrever trabalhos na construção de medidas de traços latentes, desenvolvendo para isso o método “Lei dos Julgamentos Comparativos”, um dos importantes precursores probabilísticos da TRI. Com isso surgiram dois mecanismos de respostas utilizados em construção de escalas, reconhecidos atualmente como modelos acumulativos e de desdobramento (THURSTONE, 1994).

Outros precursores realizaram importantes contribuições nessa frente. São citados por Pasquali e Primi (2003); Richardson (1936) na perspectiva da comparação de parâmetros obtidos pela TCM; Tucker

(1946) sendo o primeiro a explorar a Curva Característica do Item (CCI); e Lazarsfeld (1950) que definiu o conceito de traço latente, embora ainda restrito a medidas de atitudes. Todas essas pesquisas foram substanciais para elaboração de uma proposição mais sofisticada que posteriormente seria denominada de TRI.

Foi então na década de 1950 que Frederic Lord desenvolveu o modelo unidimensional de natureza acumulativa para respostas dicotômicas, considerando inicialmente dois parâmetros: discriminação e dificuldade, porém o próprio autor sentiu a necessidade de incorporar mais tarde um terceiro parâmetro que se propusesse a resolver o problema do acerto casual. Seus estudos iniciais permitiram estimar parâmetros para os itens utilizando o modelo de ogiva normal (distribuição normal), todavia alguns anos depois Birnbaum (1957) propôs a substituição para a função logística, o que matematicamente era mais conveniente e o tratamento matemático com maior facilidade de compreensão (PASQUALI e PRIMI, 2003).

Diversos outros autores continuaram e continuam os estudos em TRI até hoje, elencando como marcos históricos: Samejima (1969) na elaboração de modelos para averiguar respostas politômicas; Bock (1972) descrevendo o modelo politômico de respostas nominais; Andrich (1978) que avançou com o modelo de Samejima, porém com categorias ordinais politômicas; Muraki (1992) com a proposição do modelo de crédito parcial generalizado; e Andrade, Tavares e Valle (2000) como uma das principais referências que desvelam tópicos importantes para construção de escalas de proficiência.

No Brasil, o primeiro registro de aplicação da TRI foi pela aplicação da análise dos resultados no vestibular de 1993 da Universidade Federal do Rio Grande do Sul (UFRGS), conforme apresentado por Albanese, Camey e Marcondes Filho (1993). Desde então, órgãos governamentais utilizam-na como estratégia de análise de resultados do desempenho de estudantes (principalmente no âmbito da educação).

Após essa retomada na linha do tempo, precisa-se delimitar o entendimento de TRI. Trata-se, portanto, de uma teoria do traço latente e de um grupo de modelos matemáticos que estão relacionados a variáveis observáveis (como itens de um teste) ou não observáveis (comportamentos efetuados pelo sujeito), por meio de um estímulo (item) que é solicitado a ser respondido pelo ser avaliado. Nesse sentido, a resposta emitida dependerá do nível de habilidade do indivíduo, onde a aptidão é a causa e a resposta o efeito (PASQUALI e PRIMI, 2003).

Na esfera educacional, relativo à medida de proficiência, Andrade, Tavares e Valle (2000) apresentam-na como um conjunto de modelos

matemáticos que representa a probabilidade de um indivíduo responder um item de acordo com os seus parâmetros e traço latente. São diretamente proporcionais, ou seja, quanto maior for a presença da característica mensurada, maior a probabilidade de responder corretamente ao item correspondente. Corroborando, Castro, Trentini e Riboldi (2010) explicam que utilizar a TRI é um “procedimento mais sofisticado”, pois além de classificar os sujeitos, traz informações do teste como um todo, principalmente a particularidade de cada item. Podem ser calculados os parâmetros para cada critério de acordo com a sua importância e significância.

Como o traço latente é o elemento basal nessa discussão, Pasquali e Primi (2003) amplia esse entendimento para constructos psicológicos que possam ser avaliados por modelos da TRI, como: variável hipotética, fator, conceito, aptidão, atitude, componente cognitivo e traço da personalidade. Para a dimensão educação, Vargas (2007) define como habilidade ou proficiência, sendo que esta será mensurada de acordo com as respostas certas ou erradas a ser averiguada por meio de um grupo de itens que satisfaçam os critérios de validade e fidedignidade que serão explanados mais adiante na subseção 2.3.5.

Embora a TRI tenha trazido inúmeros avanços no entendimento de psicometria e medidas de traços latentes, não se deve abandonar as métricas da TCT, visto que ambas não são excludentes, mas sim complementares. O tópico a seguir fará uma análise comparativa entre as duas teorias.

2.3.1 TRI x TCT

Em uma visão macro, a TCT busca explicar o comportamento do resultado final de um teste em que uma população foi investigada, se preocupando com o desempenho do todo, do escore total (PASQUALI, 2009). Continuando, Klein (2009) afirma que o conjunto de questões e o grupo dos indivíduos é o mais importante aqui, considerando as análises e interpretações limitadas à coletividade. A comparabilidade, por exemplo, somente tem sentido quando o mesmo grupo é submetido a prova idêntica (ou paralela), se mostrando bastante inviável e ineficaz na prática. Aqui iniciam-se alguns obstáculos conforme identificados pelo autor, como a necessidade dos avaliados responderem a um amplo conjunto de questões para mensurar sua proficiência e inaptidão para avaliar a série temporal, pois as provas não podem ser as mesmas ao longo dos anos.

Já para a TRI, Andrade, Tavares e Valle (2000) explicam que o mais importante é o item e não o resultado global do teste como um todo. Isso permite realizar paridade entre os resultados dos indivíduos em testes diferentes ou ao longo dos anos, dado que uma parcela da prova é comum e possui parâmetros conhecidos para balizamento dos demais. Embretson e Reise (2000) acrescenta que itens e indivíduos podem ser posicionados na mesma escala, permitindo realizar analogias entre a característica exigida no teste, bem como a interpretação da evolução do traço latente no crescente da escala.

A TRI emerge, portanto, como uma importante alternativa para superar entraves da TCT, descritos por Hambleton, Swaminathan e Rogers (1991) como:

- a) Dimensão e alcance da parametrização de itens, pois na teoria clássica não há segurança em termos de representatividade nas amostragens quando comparadas a população, sendo que na TRI estes parâmetros passam a ser independentes da amostra (*not group-dependent*).
- b) A dependência do teste utilizado na medição da característica do traço latente, visto que somente com a TRI esta passa a ser independente do instrumento (*not test-dependent*). Logo, se estiver medindo o mesmo traço latente irá apresentar o mesmo nível de aptidão do indivíduo, desde que respeitado o erro da medida.
- c) A precisão, também reconhecida como fidedignidade, já que na medida clássica era postulado o erro da medida como idêntico entre todos os sujeitos avaliados, o que estatisticamente é praticamente improvável, enquanto na teoria moderna essa suposição não necessariamente se faz igual para todos os indivíduos da amostra.
- d) A análise do item na TCT depende do conjunto dos demais em função do escore total, enquanto na TRI é possível particularizar e avaliar cada etapa do instrumento de medida de forma independente. Dessa forma é possível contemplar diferentes graus de dificuldade no mesmo teste, além de medir os vários níveis de proficiência e permitir a realização de testes adaptativos e/ou otimizados.

Para Fletcher (1994), o mais relevante em relação à “nova” teoria é:

[...] a promessa de fornecer medidas invariantes do desempenho cognitivo, que não dependem dos itens que compõem a prova ou das pessoas investigadas na amostra. [...] A calibração fornece a cada item parâmetros que caracterizam suas qualidades técnicas, independentes da população investigada. [...] Sendo invariantes, eles não dependem da amostra selecionada para fins de calibração. Sendo invariantes, podem ser aplicados a qualquer outra população, proporcionando resultados na mesma escala de habilidade. (FLETCHER, 1994, p. 24).

Para fins didáticos, elaborou-se um quadro resumo com as principais características mapeadas pelos autores pesquisados entre a TCT e TRI:

Quadro 1 - Comparação entre a TCT e TRI

Característica	TCT	TRI
Construção do teste	Fidedignidade e consistência interna é o elemento principal	A função de informação do teste é o elemento principal
Instrumento de avaliação	Cada instrumento produz um resultado diferente de acordo com o nível de proficiência.	Independente do instrumento produzirá o mesmo resultado para o mesmo nível de proficiência.
Foco de análise dos Itens	São analisados conjuntamente, sendo o resultado de um impactado diretamente pelo outro.	São analisados de forma separada e independente do resultado dos demais.
Parametrização da Dificuldade e Discriminação	Dependente da amostra utilizada.	Independente da amostra que foi utilizada.
Acerto ao acaso	Não é possível mensurar com os recursos matemáticos.	É possível parametrizar conforme modelos específicos da TRI.
Amplitude do traço latente	São centradas nos sujeitos com traço latente mediano, não alcançando todos os níveis de desempenho.	Permite avaliação do traço latente em diferentes níveis, alcançado os com menores e maiores desempenho.
Erro da medida	Suposição de que são iguais para todos os sujeitos.	Não necessariamente são iguais para todos os sujeitos.

Fonte: elaborado pelo autor

Para ampliar a compreensão sobre o avanço da TRI em comparação à TCT, detalha-se na sequência os principais pressupostos epistemológicos que fundamentam essa teoria e promovem um aprofundamento na análise do traço latente dos sujeitos.

2.3.2 Pressupostos epistemológico da TRI

A conceituação escolhida para compreensão da TRI é a mesma apresentada por Andrade, Tavares e Valle (2000), sendo um grupo de modelos matemáticos que expressam a probabilidade de um sujeito responder corretamente um determinado item em função do nível de traço latente do respondente. Essa equação é formulada de modo que, quanto maior a habilidade avaliada, maior probabilidade de acerto no item pelo estudante.

Para ampliar a compreensão das características fundamentais que embasam essa teoria, alguns axiomas – pressupostos epistemológicos – são necessários de serem estabelecidos em relação as respostas dos itens e da aptidão dos seres avaliados. Pasquali e Primi (2003) introduzem dois postulados elementares: a) o desempenho do sujeito exposto a um determinado item pode ser predito por meio de um conjunto de fatores (traços latentes), sendo estes a causa e o desempenho o efeito; b) a relação entre desempenho e habilidade pode ser apresentada por uma formulação matemática nomeada de CCI. Lai *et al.* (2003) sintetiza também em duas suposições básicas consideradas pela maioria dos modelos de TRI: a independência local e a unidimensionalidade.

A independência local propõe que as respostas fornecidas sejam estaticamente independentes, significando dizer que a probabilidade do indivíduo acertar um determinado item não possui relação direta com os demais e depende exclusivamente da capacidade latente do avaliado (KOLEN e BRENAN, 2010). Também pode ser conhecida como princípio da invariância, pois os parâmetros dos itens independem do traço latente do respondente e vice-versa (ANDRADE, TAVARES e VALLE, 2000).

Já a unidimensionalidade defende que, embora seja compreensível que o desempenho humano seja multifacetado, o conjunto de itens avalia um único traço latente preponderante dentre os vários que influenciam o comportamento. Andrade, Tavares e Valle (2000) refletem que a unidimensionalidade já implica na independência local e, portanto, tem-se uma e não duas suposições a serem verificadas.

A dimensionalidade do teste, nessa vertente, pode ser de um único fator latente (unidimensional) ou mais de um (multidimensional) e quanto mais unidimensional for a composição do constructo, menos dúbios serão os resultados e interpretações. Neste caso haverá um único teta (θ) para cada examinado, ou seja, um determinado nível de traço latente e que quaisquer outros fatores que influenciam sua resposta são considerados como aleatórios ou erros isolados que não afetam os demais (DEMARS, 2010). Edelen e Reeve (2007) orientam que havendo dúvidas quanto a dimensionalidade do teste deve-se avaliar o comportamento dos parâmetros de discriminação dos itens, uma vez que inclinação baixa pode indicar problemas nesse aspecto.

Outros métodos podem ser empregados para verificar a unidimensionalidade. O mais difundido é a análise fatorial e permite identificar o número mínimo de fatores que podem ser explicados de acordo com a quantidade da variância observada (EMBRETSON e REISE, 2000). Segundo os autores, não existe unidimensionalidade absoluta, portanto, deve ser medida em graus. O índice é expresso pelo quociente entre a variância explicada pelo primeiro fator e a pelo segundo, sendo que quanto maior for essa explicação maior será a unidimensionalidade. Primi e Almeida (1998) esclarecem ainda que o raciocínio é simples, dado que um teste será unidimensional quando for submetido a análise fatorial e emergir um único fator predominante.

Os pressupostos aqui descritos sustentam a grande parte dos modelos de TRI que são utilizados para realizar a parametrização dos itens e estimação das proficiências dos indivíduos. Suas variações ocorrem em consonância com as características do teste e da população avaliada, influenciando a escolha do método mais apropriado. A próxima seção propõe-se em apresentar estes atributos, assim como a escolha do modelo que será aplicado para análise e tratamento dos resultados.

2.3.3 Modelos de TRI

Os modelos matemáticos construídos a partir da TRI expressam, sobretudo, a relação entre o traço latente medido pelo instrumento e a resposta fornecida ao item. Conforme apresentado por Andrade, Tavares e Valle (2000), estes variam segundo três fatores:

- natureza do item (dicotômicos ou politômicos);
- número de populações envolvidas (uma ou mais); e
- quantidade de traços latentes medidos (unidimensional ou multidimensional).

Quanto a natureza do item, Demars (2010) traz um esclarecimento adicional sobre os dois tipos de comportamentos: dicotômicos – são aqueles quando há duas categorias possíveis, isto é, certo ou errado, concordo ou discordo, sim ou não; e politômicos – quando há mais de duas categorias possíveis de serem contempladas, como a classificação gradual da escala Likert, por exemplo.

Outro fator a ser analisado na escolha do modelo é o seu caráter quanto ao processo de resposta, se cumulativo ou de desdobramento. No primeiro, a probabilidade de acerto em um determinado item cresce conforme aumenta o traço latente. No segundo padrão, a probabilidade de acertar está relacionada ao valor da habilidade do sujeito estar próximo do valor do atributo mensurado (BORTOLOTTI; ANDRADE, 2007). Especificamente sobre os modelos cumulativos, os autores destacam o modelo logístico de um parâmetro (RASCH, 1960), modelo logístico de dois parâmetros (LORD, 1952) e modelo logístico de três parâmetros (LORD, 1980). Já para os modelos acumulativos politômicos, são sinalizados o modelo de resposta nominal (BOCK, 1972), o modelo de resposta gradual (SAMEJIMA, 1969) e o modelo de crédito parcial generalizado (MURAKI, 1992). Na sequência é possível detalhar um pouco mais cada um destes modelos:

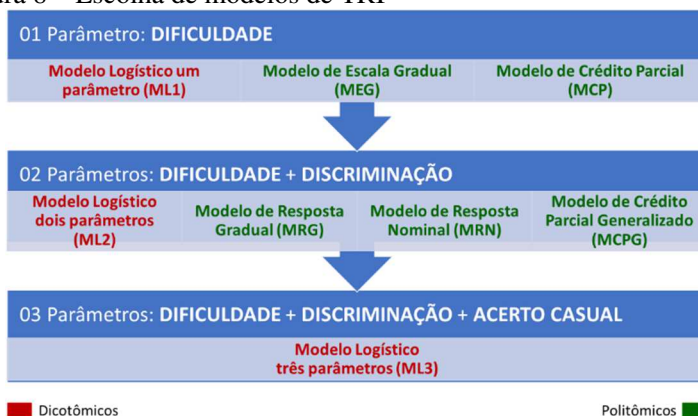
- Modelo (dicotômico) logístico de um parâmetro (ML1P): foi codificado pela primeira vez pelo matemático Rasch (1960) e expressa a probabilidade de se obter uma resposta assertiva em função da diferença entre o sujeito e a dificuldade do item, sendo este o parâmetro exclusivo de análise.
- Modelo (dicotômico) logístico de dois parâmetros (ML2P): proposto pela primeira vez por Lord (1952) e aperfeiçoado mais tarde por Birnbaum (1968). Foi idealizado inicialmente pela distribuição gaussiana padronizada e posterior transposta para a função logística. Neste modelo é acrescido o parâmetro de discriminação, permitindo assim verificar a inclinação de cada item na CCI.
- Modelo (dicotômico) logístico de três parâmetros (ML3P): pensando em testes em que o indivíduo poderia acertar um item sem necessariamente possuir o traço latente adequado, Hambleton (1991) propôs o terceiro parâmetro “acerto ao acaso”, adicionando-o no modelo com os dois parâmetros já estudado por Lord.
- Modelo (politômico) de respostas graduais (MRG): desenvolvido por Samejima (1969), foi o primeiro modelo a

pensar em categorias ordinais politômicas, considerado uma extensão do ML2P, onde não se pretende medir o certo ou errado, mas sim a gradação das respostas. Seu método baseia-se na probabilidade de um sujeito responder uma categoria ou superior a ela quando apresenta um determinado valor de traço latente.

- Modelo (politômico) de crédito parcial (MCP): trabalho por Masters (1982), é aplicado para itens ordinais utilizando apenas como parâmetro a dificuldade. É um modelo mais direto e mede apenas uma etapa da probabilidade de um indivíduo responder uma determinada categoria (diferente do MRG),
- Modelo (politômico) de crédito parcial generalizado (MCPG): estudado por Muraki (1992) é uma generalização do modelo anterior, e, portanto, aplicado para itens ordinais, porém com a inclusão do parâmetro da discriminação.
- Modelo (politômico) de respostas nominais (MRN): introduzido por Bock (1972), é utilizado para itens cujas alternativas não apresentam uma ordem determinada previamente, podendo apresentar conjuntos puramente nominais ou para validar o ordenamento esperado entre alternativas.
- Modelo politômico de escala gradual (MEG): descrito por Andrich (1978), se ajusta bem aos itens com categorias ordinais politômicas, porém diverge do MRG por considerar que os escores são espaçados igualmente entre as alternativas de resposta.

Visando esquematizar os modelos apresentados, a Figura 8 auxilia na escolha do modelo conforme número de parâmetros a ser considerado e o tipo de resposta ao item, se dicotômico ou politômico.

Figura 8 – Escolha de modelos de TRI



Fonte: elaborado pelo autor

Em função das características desse estudo, não foram descritos modelos de desdobramento, limitando-se apenas a apresentar a expressão matemática do ML2P, representada pela equação (1) devido sua escolha para tratamento dos dados dessa pesquisa:

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

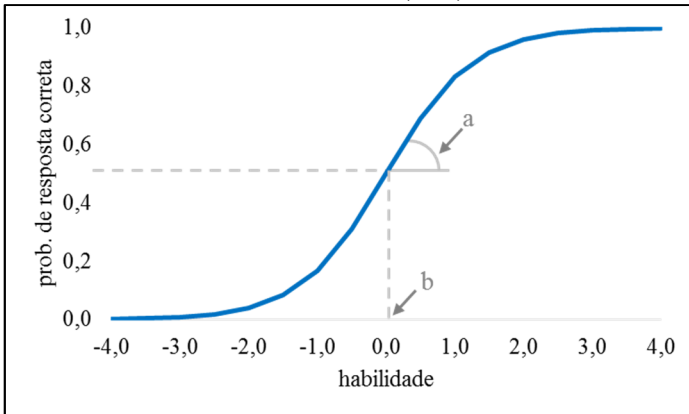
Onde:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j responde incorretamente ao item i .
- θ_j representa a habilidade (traço latente) do j -ésimo indivíduo.
- $P(U_{ij} = 1 | \theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item (FRI).
- a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da CCI no ponto b_i .
- b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade.
- D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultados semelhantes ao da função ogiva normal.

Muniz (1990) relata que uma importante relação entre os parâmetros θ_j e a_i é que a discriminação varia conforme o nível de traço latente do sujeito. Nos casos onde, por exemplo, os parâmetros incidem é possível identificar o ponto de inflexão da curva, isto é, onde muda a direção (inclinação) em relação ao eixo das ordenadas. Pasquali e Primi (2003) explicam que o parâmetro de dificuldade deste modelo é o ponto na escala de aptidão que a probabilidade de resposta correta é de 50% (0,50), podendo assumir valores que vão de $-\infty$ a $+\infty$, mas que na prática vai de -3 a +3.

Graficamente, os parâmetros do modelo podem ser expressos por meio de uma CCI, como representado no Gráfico 2.

Gráfico 2 - Curva Característica do Item (CCI)



Fonte: elaborado pelo autor

Interpretando, o eixo x representa a dificuldade e apresenta a mesma escala do nível de habilidade. Isto significa dizer que quanto maior seu valor, maior o nível de proficiência demandado por aquele item. Sendo assim, o parâmetro “b” representa o valor onde a probabilidade de resposta assertiva é de 0,50. O parâmetro “a” indica a discriminação e é proporcional à derivada da tangente da curva no ponto de inflexão (aponta a inclinação da curva). Isso significa dizer que quanto maior for este parâmetro, mais inclinada será a curva e, conseqüentemente, mais estreito será seu intervalo de discriminação.

Após a calibração dos itens, precisa-se realizar uma análise crítica sobre a qualidade dos parâmetros para possibilitar seu posicionamento na escala de traço latente, passo esse que se delinea na seção subsequente.

2.3.4 Construção da escala do traço latente

Retornando ao entendimento de medida do traço latente, Nunnally (1978) explica que os números são referências utilizadas para expressar a quantidade de uma capacidade, por exemplo, preocupando-se justamente com o quanto de um determinado atributo está presente no sujeito. Logo, a utilidade de um método de medição passa por: a) posicionamento dos dados obtidos ajustados em um modelo matemático; b) confiabilidade da medida; c) validade da medida; e d) a inter-relação do método com outras medidas científicas já estruturadas.

Estas medidas podem assumir formas diferentes em relação às propriedades do objeto, sendo que Pasquali e Primi (2003) nos introduzem três delas:

- medida fundamental: aplicada a atributos naturais e que possam ser observados de forma direta e fundamental, quando o instrumento possui a mesma qualidade daquilo que está sendo mensurado;
- medida derivada: ocorre de forma indireta por meio do estabelecimento da interação entre elementos observáveis, onde se faz inferência sobre sua relação com o atributo do objeto medido; e
- medida por teoria: atribuída para traços latentes que não são observados diretamente, aferidos apenas com base em leis e teorias científicas.

Esta última, segundo o autor, é a mais aplicada em ciências sociais por tratar de artefatos intangíveis, pois possibilita o tratamento científico da aptidão avaliada por meio de variáveis observáveis. Nesse sentido, a proficiência – objeto central deste estudo – pode ser medida em estudantes que demonstram diferentes níveis de desempenho ao ser convidados a solucionar itens (questionamentos).

Em outra referência, Pasquali (1999) apresenta um modelo para construção de escalas psicológicas que pode ser aplicado para medir proficiência, devendo ser observadas três etapas. A primeira trata de procedimentos teóricos, entendidos como uma prévia à medição, quando são delimitadas as características do traço latente, realizada a definição constitutiva e operacional, a elaboração dos itens e validação dos conteúdos. Na prática, como o atributo não pode ser observado diretamente, estabelecem-se as dimensões do constructo, isto é, o que de fato se pretende avaliar, passando necessariamente pela análise de juízes

(especialistas da área do conhecimento avaliado) quanto a coerência e validade. A segunda refere-se aos procedimentos empíricos, constituindo uma amostra plausível para aplicação do teste, ou seja, é o momento da coleta de dados. A terceira e última é composta por procedimentos analíticos (ou estatísticos), configurando-se pela determinação da dimensionalidade, análise da parametrização dos itens e onde se constrói propriamente a escala.

Especificamente sobre a escala de medida, pode-se sintetizar em três ações fundamentais: 1) calcular estatísticas descritivas de cada item (percentual de acerto e correlação bisserial, por exemplo); 2) calibrar os itens, estimando seus parâmetros (dificuldade e discriminação); e 3) determinar o nível de traço latente dos indivíduos avaliados (quantificação) conforme o posicionamento na escala (TEZZA e BORNIA, 2009).

Dois pontos são chaves aqui: estimar os parâmetros dos itens e a proficiência dos sujeitos avaliados. No primeiro caso, há dois principais métodos empregados: Máxima Verossimilhança Marginal e Métodos Bayesianos. As principais características de cada um deles são descritas por Andrade, Tavares e Valle (2000):

- Máxima Verossimilhança Marginal (MVM): seu principal elemento é a propriedade assintótica, isso significa que a estimativa pode ser realizada por métodos convencionais. Entretanto, não se pode estimar parâmetro para acerto ou erro total, pois esta função não possui pontos máximos e mínimos. Demandam bastante recursos computacionais para cálculo, necessitam uma distribuição para o parâmetro θ e apresentam dificuldades na estimação de acerto ao acaso.
- Bayesiano: podem ser definidos para qualquer padrão de resposta e permitem a estimação das habilidades posteriormente a calibração dos itens. Computacionalmente são mais trabalhosos, até porque necessitam estabelecer uma distribuição a priori, considerando variáveis aleatórias para todos os parâmetros dos itens.

Na sequência os autores continuam explanando sobre as principais diferenças entre os processos de estimação para o traço latente (proficiência):

- Máxima Verossimilhança (MV): ajusta-se bem para testes com grande número de itens, produzindo estimadores não

viciados. Contudo, não está definido para alguns tipos de padrões de resposta, como as negativas ou todas as positivas.

- Bayesiano (esperança a posteriori – EAP): é definido para todos os tipos de respostas, entendido como o procedimento com o menor erro médio, sem a necessidade de utilização de métodos iterativos. Em contrapartida, produz estimadores viciados e demandam cálculos mais avançados que o método MV, além da necessidade de uma distribuição prévia para θ .
- Bayesiano (moda a posteriori – MAP): segue as mesmas características técnicas do método EAP, porém apresenta valor do erro médio ligeiramente superior.

Também não se pode abrir mão do pré-teste dos instrumentos de avaliação para verificar a adequação em relação as características dos sujeitos avaliados. Nesse ínterim, Vianna (2005) sinaliza um ponto nevrálgico em relação à vulnerabilidade da amostra, contemplando um número suficiente que permita a calibração dos parâmetros de modo estatisticamente viáveis.

Leia-se amostra, dentro desse escopo, como um subconjunto da população que será selecionada para análise do objeto de estudo e que a partir desse extrato será possível fazer generalizações para toda a população. Existem inúmeras técnicas para seleção deste grupo, porém vamos nos ater apenas a amostragem aleatória simples – utilizada nessa pesquisa – quando cada indivíduo possui a mesma chance de ser escolhido para integrar a amostra, utilizando-se de métodos randômicos para esta finalidade (MILONE, 2004).

A quantidade de elementos que participará desse extrato dependerá do número de parâmetros a serem estimados pelo modelo e complexidade do constructo avaliado. Embretson e Reise (2000) indicam que amostras a partir de 250 sujeitos já são confiáveis, porém particularizam e sugerem estudos com dados reais. Orlando e Marshall (2002) afirmam que até mesmo 200 observações já são suficientes para estimação. Porém no estudo empírico realizado por Valentini e Hauck Filho (2013) fica evidenciado que para o modelo logístico de três parâmetros (ML3P), a partir de 500 observações já não há mais variabilidade comparado com o valor real da população, apresentando alta precisão. Considerando que este estudo possui três parâmetros e, por tanto, maior complexidade, infere-se que esse número também possa ser utilizado para o ML2P.

Definida a amostra, faz-se necessário empregar recursos computacionais específicos para os métodos de estimação. Em diversas literaturas pesquisadas é possível encontrar alguns softwares com essa

finalidade, como o BILOG-G, BILOG-MG, MULTILOG e PARSCALE, por exemplo. Estes assumem que os respondentes representam uma amostra aleatória, podendo apresentar uma distribuição normal ou distribuição discreta arbitrariamente especificada pelo usuário, ou até mesmo uma distribuição empírica (ANDRADE, TAVARES e VALLE, 2000). O BILOG-MG (ZIMOWSKI *et al.*, 1996), escolhido como suporte tecnológico nessa pesquisa, permite a estimação dos modelos logísticos de um, dois e três parâmetros e são ajustados para dados dicotomizados de modo satisfatório.

Em posse dos parâmetros calculados pelos recursos computacionais, realiza-se a ancoragem, processo esse que envolve cálculos estatísticos na identificação dos itens que discriminam os respondentes conforme o nível do traço latente. Nesse entendimento, define-se a métrica da medida de posição e da medida de dispersão, base essa para a construção e interpretação da escala. Computacionalmente, os programas padronizam em (0, 1), porém didaticamente sua compreensão é de maior complexidade, podendo ser aplicadas transformações lineares para apresentação dos resultados. É o exemplo do SAEB/Prova Brasil que utiliza a escala (250, 50), neste contexto a média é 250 pontos e o desvio padrão 50 pontos (ANDRADE, TAVARES e VALLE, 2000).

Esta ancoragem ocorre pela definição de níveis âncora, isto é, pontos selecionados para análise da escala de proficiência a serem interpretados pedagogicamente, e itens âncoras que segundo Andrade, Tavares e Valle (2000) são aqueles que satisfaçam as seguintes condições:

1. $P(U = 1 | \theta = Z) \geq 0,65$; e
2. $P(U = 1 | \theta = Y) < 0,50$; e
3. $P(U = 1 | \theta = Z) - P(U = 1 | \theta = Y) \geq 0,30$.

Desta forma, para ser esta considerado âncora, este precisaria ser respondida como “Sim” por grande proporção de alunos (pelo menos 65%) em um determinado nível da escala e por uma proporção menor de estudantes (no máximo 50%) no nível imediatamente anterior. Também deveria ter uma diferença entre as proporções nesses níveis de, pelo menos, 30%.

Sequencialmente, realiza-se a interpretação da escala na perspectiva da aptidão que está sendo medida. Para os autores, é basilar descrever para cada nível da escala sua tradução do ponto de vista técnico, sendo que isto é possível por meio do posicionamento dos itens na escala e análise dos especialistas com domínio no assunto. Hambleton (2000)

contextualiza para a área educacional, informando que a análise pedagógica da escala possibilitou avanços no acompanhamento do rendimento escolar, seja na avaliação temporal entre uma série ou outra, seja na comparação das habilidades em diferentes grupos de estudantes. Desse modo, a escala retrata o nível de aprendizagem dos estudantes no momento do teste, identificando aquilo que foi aprendido e aquilo que ainda não é da proficiência do aluno.

Mas podemos assegurar que a escala de fato mede a proficiência? Será que o teste realizado possui credibilidade para identificar os níveis de desempenho dos estudantes? Para isso, deve-se verificar a validade e fidedignidade do teste, tópico este que será abordado na sequência.

2.3.5 Validade e fidedignidade do teste

O conjunto de itens que compõem o instrumento de avaliação deve apresentar dois fundamentos essenciais para credibilidade do teste: validade e fidedignidade (PASQUALI, 1997). De modo geral, a primeira refere-se ao fato de o teste estar relacionado diretamente com aquilo que se deseja medir e a segunda com a característica da confiança de alcançar os mesmos resultados em aplicações futuras, isso é, com o menor erro de medida possível.

Na dimensão validade, busca-se verificar se a prova avalia o que supostamente deveria medir, isto é, se o comportamento dos sujeitos por meio das respostas está aferindo o próprio traço latente. Desse modo, validar passa por uma testagem empírica da verificação da hipótese, lançada a priori sobre uma teoria previamente elaborada do constructo a ser avaliado (PASQUALI, 2009). São três os tipos de validade: conteúdo, critério e constructo.

A validade de conteúdo verifica o grau em que o instrumento evidencia o domínio do que se pretende medir. Nas palavras de Richardson (1989, p. 139) “o conteúdo do instrumento (as perguntas ou itens) são amostras de diferentes situações, e o grau em que os itens representam essas situações é denominado validade de conteúdo”. Por sua vez, a validade de critério consiste na condição do quanto o teste consegue prever um desempenho específico de um sujeito, devendo ser comparada por meio de técnicas que independam do próprio teste. Pode ser realizada de modo preditivo ou concorrente com a aplicação do instrumento, porém a definição de um critério válido se mostra bastante desafiador, recorrendo-se mais facilmente a validade de constructo (PASQUALI, 2004).

Finalmente, a validade do constructo já teve diversas conotações (intrínseca, fatorial e aparente) e sua definição é expressada por Nunnally (1978) como a extensão para qual o conjunto de itens mede um traço latente teórico, averiguando a legitimidade da hipótese do constructo avaliado. Pasquali (2004) afirma que este tipo pode ser trabalhado sobre diferentes prismas, como a análise da representação comportamental do constructo, análise por hipótese e a curva da informação da TRI.

Já a fidedignidade é tratada como a característica de um teste medir com o menor erro possível, em outras palavras, com precisão. Medir nesse intuito está relacionado a submeter o mesmo teste para indivíduos com distintas situações ou aplicar testes equivalentes que averiguem os mesmos indivíduos, devendo, portanto, produzir resultados semelhantes. Nesse contexto, representa a verificação se as métricas são passíveis de repetição, se tendem a ser estáveis mesmo sobre uma variabilidade de condições e se produzem resultados análogos (NUNALLY, 1978; PASQUALI, 2004).

Existem dois métodos prioritários para definir a precisão de um teste, conforme indica Pasquali (2009): a correlação e a análise da consistência interna. A primeira é utilizada nos casos de testes e retestes, quando há dois resultados para duas ocasiões distintas, verificando-se a correlação bivariada entre as duas pontuações do mesmo sujeito. Para a consistência interna existe uma série de procedimentos estatísticos e os mais frequentes são as técnicas de coeficiente alfa (α). Para essas situações há somente a aplicação de um único teste e entende-se que, se os itens covariam numa dada situação, também irão se estender em qualquer ocasião de uso do teste.

O alfa de Cronbach, principal métrica desse tipo de coeficiente, pode assumir valores entre 0 e 1, entretanto apenas os testes que apresentam resultados superior a 0,70 são tidos com confiabilidade apropriada de acordo com o critério estabelecido pelos autores Hill e Hill (1998). Ainda sobre esta aplicação Matthiensen (2011) elucida:

O uso de medidas de confiabilidade, como o Alfa de Cronbach, não garante unidimensionalidade ao questionário, mas assume que ela existe. A unidimensionalidade é uma característica de um conjunto de indicadores que tem apenas um conceito em comum (MATTHIENSEN, 2011, p. 11-12).

Contudo, embora os coeficientes de fidedignidade sejam utilizados para comparar a confiabilidade, Anastasi e Urbina (2000) apresentam o erro padrão da medição como o mais adequado para analisar os escores individuais. Em se tratando de TRI há uma maneira de expressar a medida do teste por meio de uma Função de Informação do Teste (FIT), dependendo exclusivamente do conjunto de itens aplicados e possibilitando uma estimativa de erro padrão para cada nível da escala. Esta técnica representa graficamente quanto cada grau de habilidade pode ser estimado usando especificamente um conjunto de itens (BAKER, 2001). O autor continua explicando sobre a possibilidade de elaborar um gráfico e analisar individualmente cada item (por meio da CCI), sendo que a FIT é a soma de todos os itens.

Após pesquisar esse arcabouço teórico, desafiou-se a buscar experiências já estudadas por outros autores que possuíssem propósito similar a avaliação do domínio psicomotor da aprendizagem, principalmente relacionado à educação profissional. Foram selecionados cinco casos que trazem contribuições significativas e que passarão a ser especificados na seção seguinte.

2.3.6 TRI aplicada à avaliação da educação profissional

As experiências selecionadas para compor e ser apresentada nessa pesquisa é fruto da metodologia de RSL. Foram identificados cinco artigos que abordam a perspectiva da avaliação do domínio psicomotor associado a estudantes da educação profissional. O Quadro 2 detalha o título, nome do(s) autor(es) e ano de publicação, listados por ordem cronológica decrescente.

Quadro 2 - Experiências mapeadas na RSL

Nº	Título	Autor(es)	Ano	Periódico
01	<i>Modeling the Development of Vocational Competence: a Psychometric Model for Economic Domains</i> (Modelando o desenvolvimento da competência profissional: um modelo psicométrico para domínios econômicos)	Festner, D. Klotz, V. K Winther, E.	2015	<i>Vocations and Learning</i>

Quadro 2 - Experiências mapeadas na RSL (continuação)

Nº	Título	Autor(es)	Ano	Periódico
02	<i>Modeling the Competencies of Prospective Business and Economics Teachers: Professional Knowledge in Accounting</i> (Modelagem das competências na perspectiva dos professores de administração e economia: conhecimento profissional em contabilidade)	Schnick-Vollmer, K.; Berger, S.; Bouley, F.; Fritsch, S.; Schmitz, B.; Seifried, J.; Wuttke, E.	2015	<i>Zeitschrift für Psychologie</i>
03	<i>Modeling and Measuring the Structure of Professional Vision in Preservice Teachers</i> (Modelagem e medição da estrutura da visão profissional em professores do pré-atendimento).	Seidel, T. Stürmer, K.	2014	<i>American Educational Research Journal</i>
04	<i>Measuring nursing competencies in the operating theatre: Instrument development and psychometric analysis using Item Response Theory</i> (Medindo competências de enfermagem na sala de operação: desenvolvimento de instrumento e análises psicométricas utilizando a Teoria de Resposta ao Item)	Nicholson, P. Griffin, P. Gillis, S. Wu, M. Dunning, T.	2012	<i>Nurse Education Today</i>
05	<i>Progress Testing for Postgraduate Medical Education: a Four-Year Experiment of American College of Osteopathic Surgeons Resident Examinations</i> (Teste de progresso em pós-graduação em medicina: a experiência de quatro anos do Colégio Americano em Residência Cirúrgica de Medicina Osteopática).	Shen, L.	2000	<i>Advances in Health Sciences Education</i>

Fonte: elaborado pelo autor

O primeiro artigo e mais recente, relata a experiência no contexto da educação profissional da Alemanha. Embora o artigo tenha sido publicado em 2015, os dados foram coletados em 2013 em quatro escolas de formação profissional, selecionadas de acordo com o maior volume de matrículas no país. A experiência se baseou em criar uma escala que conseguisse avaliar o nível de competência dos estudantes de economia em dois distintos momentos: iniciantes – com até 0,5 anos de formação inicial profissional e avançados – de 1,3 a 2,8 anos de formação inicial profissional.

Como base conceitual, apropriou-se dos axiomas de Dreyfus e Dreyfus (1980) que estabeleceu dois grupos de competência: “domain-linked competence” (domínio ligado a competência) não requer conhecimentos específicos para realizar a tarefa e “domain-specific competence” (domínio específico de competência) que requer conhecimentos específicos para realizar a tarefa. Os autores elucidam que no primeiro grupo, por exemplo, estaria a competência em calcular uma taxa de câmbio, que embora seja uma atividade prevista no currículo do curso, pode ser concluída com habilidade numérica geral e não específica. Já no segundo grupo estaria a correta aplicação da teoria do contrato econômico em uma determinada situação, que requer conhecimentos específicos em economia para a resolução.

A cada um dos grupos de estudantes (iniciantes e avançados) foram aplicadas um conjunto de tarefas, sendo elaborados 46 itens diferentes, no qual 16 eram comuns aos dois grupos. Do total, 28 itens objetivavam avaliar domínios ligados a competência e 18 domínios específicos de competência. Noutra classificação: 15 itens tinham como base avaliação conceitual, a qual busca averiguar se o estudante possui conhecimento em um determinado assunto; 21 eram processuais, visando identificar a habilidade das estudantes em aplicar conhecimentos em uma determinada situação; e 10 eram interpretativos, sendo o nível mais elevado e que exige a capacidade do estudante em tomar decisões mais estratégicas a partir da interpretação.

Os itens foram elaborados em formato de situação-problema e estudo de caso, onde o respondente precisava ler e interpretar. A avaliação, portanto, era teórica e não prática. As respostas eram abertas e avaliadas por “peritos” (avaliadores) que classificavam a resposta na escala de três categorias: incorreta, parcialmente correta e correta.

Quanto aos resultados, os autores observaram que o instrumento teve boa aderência dos estudantes, pois foi utilizado como preparação para o exame final. Com base na metodologia de Rasch (1960), os itens do foram classificados em quatro níveis de progressão de aprendizagem:

nível 1 – numérico e literária, nível 2 – domínio específico de conhecimento, nível 3 – domínio específico de procedimentos e nível 4 – domínio específico de interpretação. Nos três primeiros níveis ficou evidente a progressão de aprendizagem entre os grupos dos iniciantes e avançados, demonstrando que o tempo de formação profissional contribui diretamente para a aquisição dessas competências.

Os pesquisadores apontaram ainda algumas limitações da pesquisa como a sua extensão, podendo ser reproduzida e aplicada apenas naquele momento e naquele local, não podendo ser replicada para outros grupos e países. Outra questão apontada também foi a seleção da amostra, sendo que não foi probabilística (não aleatória), sem permitir a estimação de parâmetros de grau de confiança. Todavia, ressaltam que este é um “novo campo sistemático de pesquisa” que carece de avanço na compreensão da aprendizagem profissional, alertando quanto a necessidades de investigações mais profundas.

No segundo artigo, embora o foco não seja proficiência de estudantes, a experiência é válida, pois se busca avaliar a competência dos professores. Os autores sustentam o conhecimento profissional do docente como fator crucial para o ensino de sucesso, uma vez que impacta diretamente no desempenho dos estudantes.

Amparados teoricamente por Shulman (1986), dividem o conhecimento profissional em dois grupos: conhecimento de conteúdo (CK) e conhecimento pedagógico de conteúdo (PCK). No âmbito da pesquisa, cujo foco é o estudo de contabilidade como ferramenta para as áreas de administração e economia, o CK é entendido como os conhecimentos técnicos e necessários para estruturação de instruções em sala de aula. Dentro da contabilidade teríamos três principais áreas (1) o propósito, a relevância e a base jurídica em contabilidade; (2) dupla entrada da contabilidade e (3) compra e venda. Já o PCK é percebido como tornar o conteúdo acessível aos alunos e conhecimento do pensamento do aluno (cognição). Aplicando na contabilidade temos: (1) cognição do estudante e erros típicos, (2) conhecimento das tarefas como ferramenta de ensino; e (3) conhecimento das múltiplas representações e explicações.

O instrumento foi construindo com 49 itens no total, sendo que cada participante respondia apenas 28. O teste foi realizado no formato *paper-pencil-test* (papel e caneta). Os itens podiam ser de três modos: abertos, múltipla escolha ou múltipla escolha complexa. A escala de resposta foi dicotômica: certo ou errado. A amostra contemplou 1.158 futuros professores de 24 universidades da Alemanha com programas de formação de docentes em economia e administração. Os participantes

receberam €\$ 20,00 para responderem a prova.

Os principais resultados apontam que há uma alta correlação entre os dois tipos de conhecimentos CK e PCK, sendo $R=0,92$ (muito alta). Também foi constatada diferença entre o CK de homens e mulheres, sendo que os participantes do sexo masculino obtiveram desempenho significativamente melhor do que as mulheres. Além disso, uma limitação importante apresentada foi que os itens de PCK foram construídos com situações hipotéticas e não reais, o que poderia distorcer parcialmente a escala elaborada.

Igualmente, no terceiro artigo o foco não é a avaliação do desempenho dos estudantes e sim a forma como os professores criam oportunidades de aprendizagem, uma vez que esta influencia fortemente o aprendizado do aluno. A pesquisa na formação de professores, conforme relatado no artigo, é um campo novo, pois até então as medidas eram puramente subjetivas e baseadas na titulação e em cursos realizados pelos docentes. Desse modo, surge a necessidade por um modelo de teste empírico, sendo este destacado como objeto de estudo deste artigo analisado.

O foco, portanto, é avaliar a “visão profissional” dos futuros docentes, sendo esta entendida como a capacidade de perceber e interpretar características relevantes de situações ocorridas em sala de aula. Os autores acrescentam que os professores em início de carreira encontram-se, muitas vezes, incapazes de lidar com a complexidade do ambiente escolar e apresentam dificuldade na transposição do que aprenderam na universidade para o contexto real de ensino-aprendizagem.

O constructo avaliado foi dividido em três pilares: descrever, explicar e prever situações em sala de aula. A descrição é a capacidade de diferenciar claramente os aspectos relevantes de um ensinamento notado e componente de aprendizagem (ou seja, o objetivo) sem fazer quaisquer julgamentos. A explicação é capacidade de usar o que se sabe para raciocinar sobre uma situação, esclarecer o ensino e as metas de aprendizagem, ativando o conhecimento prévio dos alunos. Já a previsão é a capacidade de prever as consequências de eventos observados em termos de aprendizagem dos alunos, aplicados à prática em sala de aula. Os autores abordam que nenhuma pesquisa foi realizada até o momento para testar empiricamente a inter-relação desses três aspectos.

O instrumento de coleta de dados foi elaborado com clipes de vídeos que representassem situações reais e relevantes de sala de aula, utilizando os critérios de autenticidade e promoção da ativação do conhecimento dos professores como parâmetros de seleção. 12 clipes

foram elencados e classificados de acordo com dois ou três dos parâmetros de “visão profissional”, validados por 40 peritos (professores) que compuseram a banca de seleção. Para cada vídeo foram desenvolvidos 36 itens com respostas na escala Likert de quatro pontos: discordo, discordo parcialmente, concordo parcialmente e concordo. Os itens eram baseados em percepções dos vídeos, sendo que as respostas dos avaliados foram comparadas com a dos peritos, que configurou o referencial de critérios. Cada indivíduo verificou 6 vídeos, com 90 minutos de duração total, sendo o formato de aplicação online.

Os resultados obtidos foram coletados em três estudos com três amostras distintas, porém com público alvo semelhante. Os dois primeiros grupos foram com estudantes no quinto semestre de formação e o último com estudantes em fase de conclusão da graduação. Os cliques foram avaliados como autênticos e relevantes pelos participantes, representando bem as três dimensões. O Modelo de Rasch mostrou maior confiabilidade do que o Modelo de Créditos Parciais, e a perspectiva tridimensional apresentou resultados mais significativos que a unidimensional e bidimensional. Foram construídas escalas para cada uma das dimensões e uma escala integradora com as três abordagens. Além disso, no terceiro grupo foram aplicadas duas vezes o mesmo instrumento, com intervalos de dias diferentes em pequenos grupos. Não houve diferenças sistemáticas entre os dois pontos de medição ao longo do tempo e os autores não apontaram limitações para a pesquisa.

O quarto artigo, está associado à área de enfermagem. Os autores buscaram elaborar um instrumento para avaliar o desempenho clínico de enfermeiros em salas de operação com base em evidências circundantes, sendo esta uma ferramenta para atestar se os enfermeiros estavam aptos a gerir e prestar cuidados de qualidade para os pacientes durante procedimentos cirúrgicos. Segundo eles, testes de múltipla escolha e testes de habilidades clínicas já foram aplicados nas escolas de enfermagem da Austrália, porém há uma necessidade e um movimento em direção a sistemas de avaliação baseados em competências. Isso se justifica porque um instrumento deste tipo consegue levar em consideração diferentes variáveis presentes no ambiente clínico que poderiam influenciar o resultado da avaliação.

Para isso, os autores propuseram um instrumento baseado em dois métodos: analítico e holístico. No primeiro buscava-se examinar aspectos específicos do desempenho em relação a um conjunto de critérios, isto é, julgava-se a performance de modo geral, mas para cada aspecto era atribuído uma pontuação separadamente. Na visão dos autores, esse método permite níveis de informações mais detalhados e possibilita um

feedback mais pontual ao avaliado. Já o holístico verificava a qualidade global do desempenho, não marcando os componentes separadamente. Apesar de prestar poucas informações, entendeu-se válido verificar se o enfermeiro havia conseguido alcançar o resultado global da atividade esperada.

Para ambos os métodos o critério de avaliação adotado era a rubrica, podendo ser entendida nas palavras dos autores como “a ferramenta de pontuação que estabelece as especificações e expectativas para uma missão”. Isto é, a qualidade requerida do desempenho que será verificada por meio de evidências. Para isso foram elaborados descritores comportamentais para cada item e codificados em 0, 1, 2, 3 ou 4, sendo o 4 o nível mais avançado e 0 atribuído quando o desempenho não pudera ser observado. Também foram aplicados itens dicotômicos, categorizados em “competente” ou “ainda não competente”.

O instrumento foi validado por 32 educadores em enfermagem que aplicaram em três enfermeiros de diferentes níveis de experiência. Foi solicitado feedback da ferramenta e preenchimento de questionário contextual para levantamento de dados demográficos. Os resultados evidenciaram que com a aplicação do modelo de Rasch (1960) a maioria dos itens apresentou intervalo de estimativas de ajustes aceitáveis, com limite definido entre 0,7 e 1,3. Os autores concluem que a evidência é necessária para reduzir a subjetividade, usando uma abordagem de lista de verificação (padrões esperados de comportamento) e que a análise por meio do modelo de Rasch é um bom caminho para a tomada de decisão em competência.

O quinto artigo e último traz a experiência do Colégio Americano de Cirurgiões Osteopáticos (ACOS) e do Conselho Nacional de Examinadores Médicos Osteopatas (NBOME), que desde 1994 aplicam testes de progressos no programa de pós-graduação em medicina. Estes possuem como objetivo fornecer informações aos coordenadores sobre o diagnóstico de aprendizagem dos residentes, sem possuir caráter de aprovação/reprovação.

O método utilizado é a aplicação de instrumentos por escrito, com periodicidade anual e itens no formato dicotômico (certo ou errado). É conduzido nos quatro anos de formação, somando um total de 300 itens por teste, distribuídos em nove temas clínicos e que contemplem todo o currículo (independente do ano que o residente esteja cursando). Outras características identificadas foram: 50 itens formam intersecção com anos anteriores ou posteriores; pessoas e itens foram posicionados na mesma escala interpretativa; e utilização do modelo Rasch, apresentando confiabilidade razoável segundo o autor.

Quanto aos principais resultados, de modo geral, os exames foram classificados como difíceis, uma vez que cada teste de progresso cobriu todo o currículo e os avaliados em níveis de formação mais baixos estavam claramente em desvantagem. O crescimento entre os anos é contínuo, porém não linear. No nível individual, os testes ajudam os coordenadores da pós-graduação a avaliar se os residentes estão cumprindo as normas e verificar o quão rápido estão se aproximando dos padrões necessários. No nível geral, serve com uma ferramenta para avaliar a eficácia de cada programa. As limitações apresentadas foram: necessidade de mais análises psicométricas junto a amostra, em especial a avaliação da validade preditiva, além da demanda pela construção de um banco de itens para sistematização do processo.

3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo serão descritos os aspectos metodológicos trabalhados no decorrer deste estudo. Inicialmente será realizada a caracterização da pesquisa e a composição da amostra selecionada para estudo. Na sequência serão apresentados o constructo utilizado para medir o domínio psicomotor, bem como o instrumento de avaliação e os procedimentos de coleta de dados. Posteriormente serão elucidados os métodos de estimação dos dados quanto aos parâmetros dos itens e proficiências, bem como as etapas de construção e interpretação da escala.

3.1 CARACTERIZAÇÃO DA PESQUISA

Não há um consenso quanto aos procedimentos de caracterização da pesquisa científica, sendo um critério do pesquisador a escolha dos metodólogos e das dimensões de análise. Gil (2008) sugere a classificação por intermédio dos objetivos e dos procedimentos técnicos utilizados no estudo. Além de se ancorar nessa perspectiva, soma-se a definição de outros autores, como Marconi e Lakatos (1999), Zanella (2006), Silva (2010) e Vergara (2013).

Inicialmente, em relação a sua natureza, esta pesquisa é do tipo teórico-aplicada, pois visa construir uma escala que permita avaliar na prática o domínio psicomotor de estudantes da educação profissional à luz da TRI. No viés teórico, motivou-se na análise de autores sobre o tema e experiências internacionais por intermédio da RSL. Já no campo aplicado, colocou-se em prática diversos procedimentos estatísticos de tratamentos dos dados que possibilitassem a criação da escala, bem como a ponderação por um grupo de especialistas técnicos na sua leitura e interpretação pedagógica.

Referente a abordagem, é predominantemente do tipo quantitativo, porém apresenta aspectos qualitativos, configurando-se como quali-quantitativa. Zanella (2006) explica que o método quantitativo está aportado na representatividade numérica e se preocupa com a medição objetiva de um fenômeno e a quantificação do seu resultado. Portanto, se propõe a delinear generalizações estudando apenas uma parcela da população (amostra) em condições estatisticamente viáveis. A razão pela escolha da priorização desta abordagem deu-se pela elevada correlação com o objetivo geral, uma vez que para estabelecer uma escala é necessário aplicar técnicas matemáticas que sejam capazes de mensurar quantitativamente o traço latente objeto de análise. Todavia, o método qualitativo demonstrou-se como um sustentáculo desta pesquisa, pois

embora não objetive a medição dos dados, preocupa-se em conhecer em profundidade a habilidade a ser avaliada, desde sua perspectiva histórica até seus pilares de fundamentação.

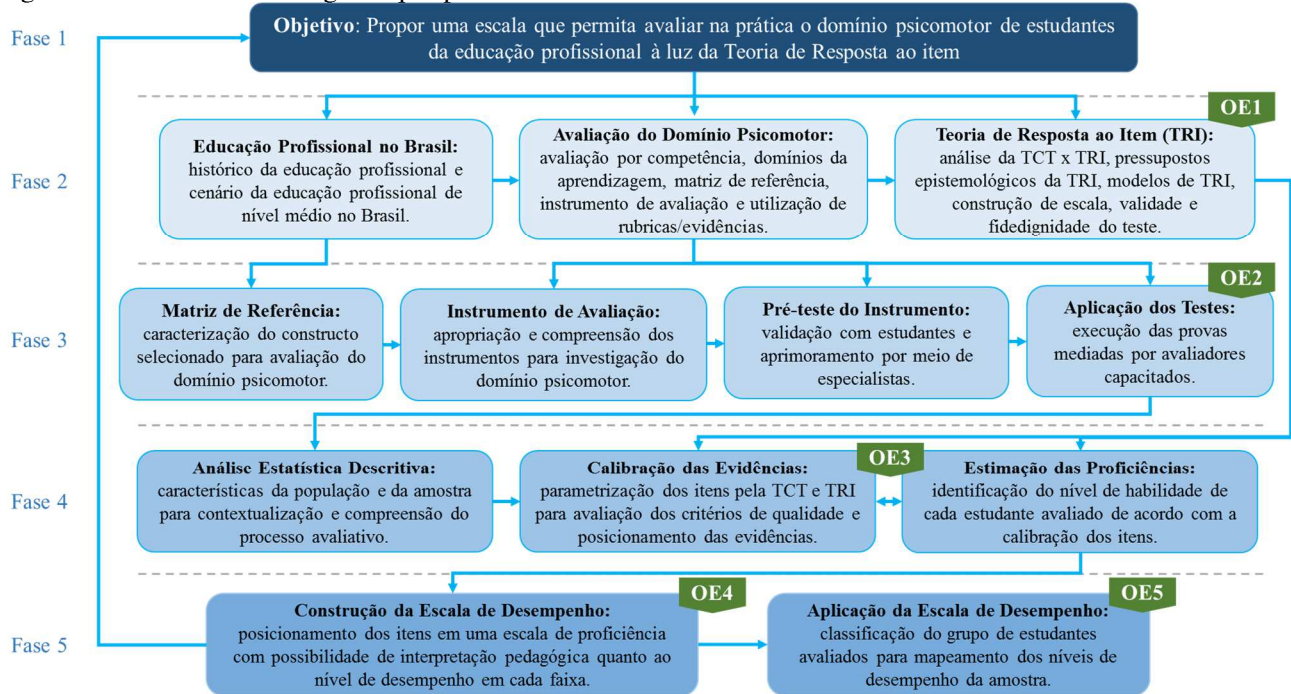
Quanto ao objetivo ou finalidade, classifica-se como exploratória e aplicada. Na visão de Marconi e Lakatos (1999) a pesquisa exploratória caracteriza-se pelo emprego do método científico para explorar o desconhecido e evoluir no conhecimento, pois o levantamento do estado da arte possibilita bases suficientes para realizar análises consistentes. Já a pesquisa aplicada diferencia-se pelo seu interesse prático na resolução de um problema decorrente de uma realidade, neste caso na investigação e mensuração da proficiência dos estudantes.

Os métodos de pesquisa, também denominados de procedimentos técnicos, podem ser tipificados como bibliográfico, experimental, *ex-post facto*, estudo de corte, levantamento, estudo de campo, estudo de caso, pesquisa ação e pesquisa participante (GIL, 2008). Foram selecionados para esta dissertação os tipos bibliográfico e levantamento. O primeiro compreendeu a consulta em literatura técnica, artigos científicos publicados em periódicos, teses e dissertações em repositórios nacionais e internacionais, além da RSL descrita na seção 3.2. Já o segundo consistiu no acompanhamento da coleta de dados referentes a amostra de estudantes que foram submetidos a avaliação prática, processo descrito mais adiante.

Por fim, quanto as técnicas de coletas de dados, pode ser segmentada em dois modos de observação: indireto e direto. A observação indireta abrangeu a análise bibliográfica e compreensão do traço latente no campo teórico, previamente ao desenvolvimento das aplicações práticas. Já na observação direta foi verificado *in loco* a aplicação dos instrumentos de coletas de dados, bem como o processo avaliativo dos estudantes da educação profissional na investigação das suas proficiências.

Além das classificações já apresentadas, essa pesquisa foi dividida em cinco fases (conforme Figura 9) para facilitar o entendimento metodológico.

Figura 9 - Fases da metodologia de pesquisa



Fonte: elaborado pelo autor

A primeira fase tinha como objetivo realizar o planejamento da pesquisa e seus principais desdobramentos, trazendo foco para o trabalho do pesquisador em relação as dimensões do estudo. Definiu-se o objetivo geral e específicos, o problema de pesquisa e sua justificativa, bem como a delimitação do tema de pesquisa. Os resultados dessa etapa estão apresentados no primeiro capítulo.

Na sequência, a segunda fase caracterizou-se pelo levantamento bibliográfico para subsidiar as discussões e o procedimentos técnicos abarcados. Inicialmente foi trazido o contexto da educação profissional no Brasil na perspectiva histórica e da habilitação técnica (formação vinculada a população). A segunda temática foi a avaliação do domínio psicomotor, construindo um olhar que passa pela avaliação por competências; os domínios da aprendizagem e suas taxonomias; a matriz de referência como uma forma de representação gráfica do constructo a ser mensurado; os tipos de instrumentos de avaliação; e o emprego de rubricas/evidências. O terceiro tópico apresentado tratava sobre a TRI como métrica de análise dos resultados para construção de escalas de proficiência, envolvendo análise comparativa entre as teorias de avaliação; os pressupostos epistemológicos e modelos; a construção de escalas; e a validade e fidedignidade do teste. Por fim, descreveu-se algumas experiências envolvendo avaliação do domínio psicomotor pesquisadas por meio da RSL. Com esta estratégia buscou-se contemplar o primeiro objetivo específico que é de “investigar experiências empregadas na avaliação do domínio psicomotor utilizando a TRI como método de análise e interpretação de resultados”. Os levantamentos dessa etapa estão descritos no segundo capítulo.

Em seguida, a terceira fase preocupou-se em desenvolver o segundo objetivo específico: “caracterizar o experimento objeto de estudo quanto a elaboração do constructo, instrumento de avaliação utilizado e processo de aplicação das provas práticas”. Dividiu-se essa fase em quatro partes, trabalho: a matriz de referência como constructo, apropriação e compreensão dos instrumentos de avaliação, descrição do processo de validação e a aplicação dos testes, suas características e abordagens na perspectiva da coleta de dados. Os resultados dessa etapa estão parcialmente distribuídos no terceiro e quarto capítulo.

A quarta fase por sua vez foi o momento de análise e tratamento dos dados e está correlacionada ao alcance do terceiro objetivo específico “realizar a estimação dos parâmetros dos itens de avaliação e proficiência dos estudantes na perspectiva do domínio psicomotor”. Foram empregados esforços em três subdivisões, sendo a análise estatística descritiva para caracterização da população e da amostra; a calibração dos

itens, calculando seus parâmetros e realizando o posicionamento deles; e a estimação das proficiências, atribuindo valor para cada nível de habilidade desempenhado. Os detalhes dessa etapa foram sistematizados no quarto capítulo.

Por fim, a quinta fase se subdividiu em dois processos. Para atender o quarto objetivo específico “elaborar escala de níveis de desempenho que permita a identificação da qualidade da proficiência do domínio psicomotor dos estudantes da educação profissional”, elaborou-se uma escala interpretativa quanto ao nível de desempenho em cada faixa que trouxesse sentido pedagógico para docentes da área avaliada. Encerrando, é realizada classificação dos estudantes de acordo com a escala proposta, visando garantir o quinto e último objetivo específico “aplicar a escala construída na amostra selecionada para diagnóstico dos níveis de desempenho dos estudantes da educação profissional”. A resolução desta etapa também está apresentada no quarto capítulo.

Nas próximas seções serão abertos os procedimentos metodológicos utilizados em cada fase da pesquisa em correspondência a Figura 9.

3.2 REVISÃO SISTEMÁTICA DA LITERATURA

Como procedimento exploratório, foi utilizada a RSL, caracterizada por ser um método que se propõe a responder uma questão específica de pesquisa, utilizando técnicas e procedimentos explícitos e organizados, possíveis de ser reproduzidos (DICKSON, 1999). Nesse sentido, foram identificados, selecionados e avaliados criticamente os estudos já elaborados e correlacionados com a problemática da dissertação, objetivando eliminar tendenciosidades na escolha das fontes analisadas.

A RSL realizada segue as orientações da publicação de Cochrane Handbook 1 (CLARKE, 2001), que determina uma sequência de sete passos: formulação da pergunta; localização e seleção das bases de dados para pesquisa; avaliação crítica dos estudos; coleta de dados; análise e apresentação dos dados; interpretação dos dados e aprimoramento e atualização da revisão.

Após a formulação da pergunta – “quais experiências de aplicação da TRI na avaliação do domínio psicomotor já foram pesquisadas e registradas cientificamente no contexto da educação profissional?” – foi escolhida como base de dados a plataforma Scopus, por ser a maior fonte referencial de literatura técnica e científica, possuir mais de 22.000 títulos de mais de 5.000 editoras. É uma base internacional e que abrange

publicações em várias áreas do conhecimento (multidisciplinar), além de possuir recursos de inteligência computacional na indexação de artigos.

Para avaliação crítica e coleta de dados, primeiramente foi inserido o termo "Item Response Theory" no programa, pesquisado apenas em títulos, abstracts e keywords, resultando em 4.994 publicações. Além disso, limitou-se a pesquisa para que as publicações possuíssem essa terminologia como palavra-chave, reduzindo para 2.266 documentos encontrados, com o objetivo de que os estudos localizados tivessem esse tema como principal. Na sequência foram inclusos descritores que remetessem a educação profissional como: "professional education", "professional training", "professional development", "vocational education" ou "work education", resultando 18 publicações. Das 18 publicações selecionadas, apenas 09 estavam disponíveis gratuitamente, sendo que as excluídas não possuíam título ou resumo que apontassem o investimento. Dessas, quatro foram descartadas por focar exclusivamente na educação básica (ensino de matemática e ciência), sendo que não atendem a pergunta de pesquisa, restando apenas cinco artigos. O quadro abaixo resume os passos utilizados na RSL.

Quadro 3 - Filtros aplicados na RSL

Filtro	Resultados	Observações
TITLE-ABS-KEY (" <i>Item Response Theory</i> ")	4.994	-
AND (LIMIT-TO (EXACTKEYWORD, " <i>Item Response Theory</i> "))	2.266	-
AND ("professional education" OR "professional training" OR "professional development" OR "vocational education" OR "work education")	18	-
Artigos disponíveis gratuitamente	09	Nove artigos excluídos por esse critério.
Excluídos artigos que não possuíam como foco a área da educação profissional	05	Quatro artigos excluídos por esse critério.
Amostra final	05	Compõe a amostra final da análise descritiva.

Fonte: elaborado pelo autor

A TRI é um tema recente e amplamente pesquisado por diferentes áreas do conhecimento. A primeira publicação encontrada foi do ano de 1981, por Wolfe, J. H. e a culminância no ano de 2015, com 532 documentos encontrados (10,7% do total). Na linha do tempo, verifica-se uma tendência de crescimento no número de estudos sobre o tema, em especial a partir de 2004, sendo que mais de 81,2% das publicações são registradas a partir deste ano. As publicações se dividem em 28 áreas do conhecimento, com destaque para a área de psicologia (47,6% do total). Há também uma pluralidade quanto à origem das publicações, com 86 países diferentes, sendo os Estados Unidos o mais representativo (53,6% do total). Os estudos se classificam em 12 tipos diferentes, sendo que 83,1% são artigos. Além do termo "Teoria de Resposta ao Item", não existe outro descritor predominante, sendo encontrado 154 palavras-chave diferentes.

Os cinco artigos selecionados por essa abordagem metodológica foram apresentados na seção 2.3.6 deste trabalho, cujo propósito está relacionado ao levantamento de experiências internacionais (ou nacionais, caso houvesse) a ser considerada como fundamentação teórica para análises deste estudo.

3.3 POPULAÇÃO E AMOSTRA DA PESQUISA

Resgatando a definição, Freund e Simon (2000) esclarecem que a população representa o conjunto de todas as observações admissíveis (ou hipoteticamente possíveis) do fenômeno que deseja se observar. Já a amostra é tida como um subconjunto da população, isto é, uma parcela selecionada por meio de técnicas e estratégias a fim de que a qualidade do todo possa ser avaliada ou julgada.

A população aqui é retratada por 5.856 estudantes do SENAI, com mais de 80% da carga horária já realizada do curso técnico em mecânica. Embora outros cursos também tenham sido submetidos aos testes práticos, optou-se por esse em particular devido a ser um dos mais expressivos (maior população), maior índice de atividades práticas realizadas ao longo da formação, alto vínculo entre o tipo de formação com o *core business* da instituição, além da tradição e perenidade na oferta ao longo das últimas décadas. Os estudantes estão distribuídos em 18 UF, cobrindo todas as regiões do país.

Devido às características operacionais, logísticas e financeiras, não era possível aplicar as provas práticas com toda a população, sendo necessário, portanto, definir uma amostra objeto de estudo. Quanto às suas tipologias, Vergara (2013) aponta que há duas classificações em se

tratando de amostragem: a probabilística e a não probabilística, sendo a primeira quando todos os sujeitos possuem algum nível de probabilidade de serem selecionados, enquanto a segunda utiliza-se outros critérios não aleatórios, seja por acessibilidade ou conveniência, por exemplo. Nesta pesquisa adotou-se o primeiro caso, utilizando-se de software para sorteio randômico dos sujeitos (SPSS17).

Para o cálculo quantitativo do número de indivíduos optou-se pela estratégia da amostragem aleatória simples aplicados em cada Unidade Operacional (UO) – nomenclatura atribuída pelo SENAI para designar escola – independente de extratos. A utilização desse artifício visa a diluição da amostra pelos estados, permitindo sua variabilidade e minimização do erro amostral. A fórmula utilizada é apresentada por Martins e Theóphilo (2009), expressa pela equação:

$$n = \frac{z^2 \cdot \sigma^2 \cdot N}{z^2 \cdot \sigma^2 + \varepsilon^2 \cdot N} \quad (2)$$

Onde:

- z é o nível de confiança, estabelecido arbitrariamente em 95%, portanto $z = 1,96$
- σ é o desvio padrão
- N é o quantitativo de alunos em cada escola do curso avaliado
- ε é a margem de erro da estimativa de proficiência.

A avaliação prática é uma segunda etapa do SAEP (programa implantado pelo SENAI), sendo precedida de um teste online, cognitivo e já parametrizado (não detalhado devido ao escopo deste estudo). Utilizou-se como referência, para estabelecimento das amostras, o desvio padrão da última aplicação disponível até então (edição 2016), além de outras premissas delineadas a seguir:

- erro amostral admitido de 30 pontos na métrica da escala de proficiência da prova online (500, 100); e
- ajuste do cálculo amostral para que as escolas com menos de dez estudantes na turma avaliada contemple todos os discentes.

Desse modo foram indicados 1.498 estudantes que compõem a amostra final deste estudo. Se aplicássemos de modo não estratificado, 361 estudantes seriam suficientes para compor a amostra, porém como cada escola busca entender seu processo avaliativo e promover ações

pontuais para melhoria do desempenho, foi necessário ampliar seu espectro em consonância com as estratégias indicadas previamente. Corresponde, portanto, a pouco mais que um quarto da população (25,6%), configurando-se como estatisticamente viável para análises e generalizações.

Outro ponto a ser considerado é que, devido às características da TRI, existe uma discussão entre os teóricos referente a uma amostra mínima para que os resultados possam ser consistentes. Nunes e Primi (2005) realizaram um estudo testando nove tipos diferentes de tamanho de amostras aplicadas sob uma mesma população, concluindo que aquelas com mais de 200 observações são confiáveis, não recomendado amostragens menores devido a instabilidades e, conseqüentemente, afetando sua confiabilidade. Como a amostra deste estudo ultrapassa um milhar de casos, entende-se como satisfatória para atender esta orientação.

Embora não esteja no mérito desta pesquisa a coleta de dados (aplicação das provas), mas sim sua caracterização para posterior tratamento e interpretação dos resultados, as próximas seções 3.4, 3.5, 3.6 e 3.7 buscam descrever desde o constructo avaliado, passando pelo instrumento elaborado para sua verificação, a pré-testagem para elevar a qualidade do teste, até mesmo o modo como são executadas na prática.

3.4 CONSTRUCTO DO DOMÍNIO PSICOMOTOR

Para a concepção do constructo do domínio psicomotor precisou-se definir qual a unidade de análise a ser avaliada elegendo as capacidades para esse papel. No entendimento da Metodologia SENAI de Educação Profissional tem-se capacidade como:

[...] o conjunto articulado e coerente de resultados de aprendizagens que um processo formativo deve garantir para que uma pessoa possa demonstrar desempenhos competentes. Sua característica fundamental é a possibilidade de ser transferível a contextos e problemas distintos daqueles que se utilizam para seu desenvolvimento. (SENAI, 2013)

Ainda nesse ponto de vista, a capacidade expressa a necessidade do aluno em “saber”, “fazer” e “ser”, podendo ser classificada como básicas, técnicas ou de gestão. A primeira está relacionada aos fundamentos técnicos e científicos a serem desenvolvidos pelo estudante

como alicerce, comportando-se de maneira universal e passível de aplicação em diferentes ocupações. Já a segunda está associada a especificidade de cada profissão, estando diretamente relacionadas ao exercício da atividade laboral. Por fim, a terceira é aquela que abrange aspectos sociais, organizativos ou metodológicos quanto à qualidade e à organização do trabalho, às relações profissionais e à condição de responder a situações novas e imprevistas.

Ao avaliar o desenvolvimento de capacidades, como preconiza a abordagem metodológica da instituição, tem-se a concepção do desenvolvimento de competências. Estas são descritas utilizando procedimento particular e são mapeadas no perfil profissional, desdobrando-se em:

- Competência geral: expressa, de forma cumulativa e sumária, as grandes funções que o egresso do curso deve desempenhar no mercado de trabalho.
- Unidades de competência: é o primeiro nível de desdobramento da competência geral, dividindo em cada uma das grandes funções, que necessariamente expressarão atividades completas, com início, meio e fim.
- Elementos de competência: é o segundo nível de aprofundamento, analisando para cada função quais as subfunções ou etapas do trabalho que somadas possibilitarão o desempenho da unidade de competência.

As subfunções ainda são esmiuçadas no desenho curricular (currículo) discriminando quais as capacidades básicas, técnicas e de gestão são demandadas, bem como os conteúdos formativos que darão suporte para que essas sejam trabalhadas. Percebe-se, portanto, que na Metodologia empregada pela instituição de ensino em foco, a competência é alcançada quando o sujeito consegue desempenhar diferentes funções e subfunções, sendo estas somente possíveis de serem desenvolvidas quando mobilizado um conjunto de capacidades (potenciais) que permitam ao estudante a plena performance em sua atividade profissional.

Esclarecidas as nomenclaturas, precisava-se de um método de representação visual que pudesse triangular a competência x capacidades x objetos de conhecimento. Devido a essas características e já utilização como constructo do processo avaliativo cognitivo, foram mantidas as matrizes de referência já empregadas, porém noutra perspectiva. A análise foi conduzida por especialistas com formação técnica na área que

realizaram novos cruzamentos conforme as capacidades que seriam priorizadas na avaliação prática. Por questões operacionais houve a necessidade de limitação de tempo da prova para que não ultrapassem uma jornada compatível com o tempo disponível em um período de aula (visto que as avaliações ocorrem dentro deste intervalo).

Como não havia muitas referências precursoras, a adaptação realizada na matriz de referência enfatizou o nível de aplicação das capacidades, haja visto que inicialmente foram construídas utilizando a Taxonomia de Bloom (domínio cognitivo). O processo baseou-se na reflexão de capacidades que pudessem ser observadas no desenvolvimento de situações-problema desafiadoras e na entrega de “produtos” que evidenciarium a mobilização das competências profissionais. O resultado deste trabalho está apresentado na Figura 10.

Figura 10 - Matriz de referência do Curso Técnico em Mecânica

MATRIZ DE REFERÊNCIA - CURSO TÉCNICO EM MECÂNICA										
COMPETÊNCIA GERAL: Atuar no desenvolvimento de projetos, aplicar processos de produção mecânica e realizar a manutenção mecânica de máquinas e equipamentos segundo normas técnicas, considerando padrões de qualidade, de saúde e segurança no trabalho, e meio ambiente.		BÁSICAS		ESPECÍFICAS						GESTÃO
		C1	C2	C3	C4	C5	C6	C7	C8	C9
		Interpretar dados e informações	Planejar etapas do processo	Especificar recursos	Definir técnicas	Controlar processo	Analisar parâmetros do processo	Avaliar processo	Aplicar procedimentos e normas técnicas	Resolver problemas
Unidades de competências	Elementos de competências									
UNIDADE DE COMPETÊNCIA 01: Atuar no desenvolvimento de projetos segundo normas técnicas, considerando padrões de qualidade de saúde e segurança no trabalho, e meio ambiente.	1.1 Especificar materiais	-	-	-	-	-	-	-	-	-
	1.2 Planejar as etapas do desenvolvimento do projeto	-	-	-	-	-	-	-	-	-
	1.3 Projetar elementos e conjuntos do projeto	C, D	-	-	C, D, I	-	-	-	C, D, I	-
UNIDADE DE COMPETÊNCIA 02: Aplicar processos de produção mecânica segundo normas técnicas, considerando padrões de qualidade, de saúde e segurança no trabalho, e meio ambiente.	2.1 Analisar parâmetros do projeto	E, J, N	C, L	C, F	-	-	-	-	-	-
	2.2 Executar o processo produtivo	-	-	-	-	C, F, L, N	-	-	C, F, H, L, N	-
	2.3 Implementar melhorias em processos mecânicos	-	-	-	-	-	-	-	-	-
	2.4 Monitorar os parâmetros de processos mecânicos	-	-	-	-	-	-	-	-	-
	2.5 Organizar os trabalhos de acordo com as etapas do processo	-	-	-	-	-	-	-	-	-
UNIDADE DE COMPETÊNCIA 03: Realizar a manutenção mecânica de máquinas e equipamentos segundo normas técnicas, considerando padrões de qualidade, de saúde e segurança no trabalho, e meio ambiente.	3.1 Avaliar a execução da manutenção	-	-	-	-	-	-	-	-	-
	3.2 Executar a manutenção	F, G, S	-	F, U	U	F	-	-	H	-
	3.3 Implementar melhorias	-	-	-	-	-	-	-	-	-
	3.4 Instalar máquinas e equipamentos	-	-	-	-	-	-	-	-	-
	3.5 Planejar a manutenção	-	-	-	-	-	-	-	-	-
Objetos de Conhecimento										
C	Interpretação de desenho técnico	I	Normas técnicas e legislações							
D	Desenho assistido por computador	J	Tecnologia de materiais							
E	Cálculo técnico aplicado	L	Processos de fabricação							
F	Metrologia	N	Tecnologia de ferramentas de corte							
G	Ajustes e tolerâncias	S	Elementos de máquinas							
H	Normas de saúde e segurança no trabalho e ambientais	U	Técnicas de manutenção (lubrificação e manutenção aplicada)							

Fonte: adaptado de SENAI (2017)

A leitura deste documento é efetuada pela triangulação das seguintes informações: a) competência geral – desmembrada nas unidades e elementos de competências dispostas na parte lateral esquerda da matriz; b) capacidades a serem avaliadas e dispostas na parte superior, ainda subdivididas em básicas, específicas e de gestão; c) e os objetos de conhecimento, inseridos na célula onde os cruzamentos de “a” e “b” são possíveis, representados por números cujo a legenda é apresentada na parte inferior. Um exemplo para deixar mais claro: o técnico em mecânica possui como uma de suas funções principais (unidade de competência 1) “atuar no desenvolvimento de projetos (...)” e sua subfunção (elemento de competência 1.3) “projetar elementos e conjuntos do projeto”. Para que essa competência seja alcançada, ele precisará desenvolver três capacidades, conforme os cruzamentos apontados na matriz: C1, C4 e C8. Especificamente sobre a C1 “interpretar dados e informações” ele ainda necessitará mobilizar dois objetos de conhecimento, C e D, que respectivamente significam “interpretação de desenho técnico” e “desenho assistido por computador”.

A matriz de referência, portanto, é o instrumento que serviu como norteador da elaboração do instrumento de avaliação, isto porque ao definir a situação-problema e as evidências que deveriam ser verificadas, necessitou-se perceber quais as capacidades priorizadas, bem como as atividades profissionais que sinalizariam o caminho das atividades simuladas a serem incorporadas no processo, além dos objetos de conhecimento que seriam mobilizados pelo estudante para realizar as tarefas demandas.

3.5 INSTRUMENTO DE AVALIAÇÃO

Um instrumento de avaliação que considere o domínio psicomotor, necessariamente, precisa estar atrelado a uma estratégia onde se possa observar a capacidade do estudante em realizar atividades profissionais na prática, ainda que em ambiente simulado. Nesse contexto, a abordagem escolhida para avaliação dos estudantes da educação profissional foi a submissão destes a situações-problema, apresentando-lhes uma circunstância real ou hipotética, significativa e correlacionada ao mundo do trabalho da ocupação avaliada. A sustentação deste designo está ancorada na Metodologia SENAI de Educação Profissional (MSEP), pois considera-a como a principal estratégia de ensino-aprendizagem da instituição.

Para definição das atividades práticas as quais os estudantes deveriam desenvolver, foram levadas em consideração as capacidades

previstas e priorizadas na matriz de referência. Orientou-se que aos autores – docentes com conhecimento técnico da área e de diferentes localidades do território nacional – selecionassem o maior número possível de capacidades da matriz de referência adequadas às condições de operacionalização, como o tempo de execução e aspectos de infraestrutura.

Foram elaborados nessa fase os seguintes documentos:

- caderno de prova: traz a situação-problema que deverá ser solucionada pelo estudante, detalhando quais as entregas e processos que deverão ser considerados como resultados do processo avaliativo;
- lista de infraestrutura: detalhamento dos equipamentos, instrumentos, materiais de escritório, ferramentas, materiais de consumo e equipamentos de segurança demandados para operacionalização da prova prática;
- manual do avaliador: descreve todos os aspectos de execução da avaliação, desde a organização dos postos de trabalho até as orientações para correção das evidências quanto aos seus atendimentos ou não; e
- lista de verificação: caracteriza-se por elencar os critérios de avaliação traduzidos em evidências, isto é, aquilo que se observará no desempenho do estudante para determinar se este é capaz ou não de realizar uma determinada atividade.

Na sequência serão desdobrados cada um destes documentos que compõem os instrumentos de avaliação.

3.5.1 Caderno de prova

Apresenta-se como o documento a ser entregue para o estudante no momento da aplicação, contendo uma ou mais situações-problema relacionadas ao curso avaliado, além de incluir todas as informações necessárias para o alcance dos resultados e desenvolvimento dos produtos esperados pelos estudantes (como gráficos, tabelas, trechos normativos e anexos com *templates* ou detalhes adicionais ao problema exposto). Nele estão descritos os resultados e entregas esperados e que irão propiciar o subsídio para produzir e/ou demonstrar as evidências de processo ou produtos resultantes da resolução.

Como diretrizes norteadoras deste processo, foram solicitados que os docentes/ elaboradores do instrumento levassem em consideração:

- os desafios escolhidos devem estar em consonância com a matriz de referência e as capacidades priorizadas para o curso;
- a necessidade de considerar como tempo máximo de execução três horas de prova, uma vez que este tempo é compatível com o horário de realização das aulas dos alunos, não provocando impactos significativos na rotina da escola;
- os aspectos de infraestrutura e realidades locais, alinhando as necessidades ao processo formativo do curso, destacando que os recursos necessários não deveriam trazer citações de modelos, marcas ou processos específicos de equipamentos e ferramentas, para que não inviabilizasse a reprodução em distintos estados; e
- a possibilidade de serem fornecidas etapas já desenvolvidas e que demandam continuidade ou finalização.

No Anexo A verifica-se um exemplo de situação-problema abordada no caderno de prova do curso técnico em mecânica, objeto de estudo desta dissertação.

3.5.2 Lista de verificação

Conforme as características operacionais da prova prática escolhida, os elaboradores elencaram os recursos necessários para aplicação das provas. Como os currículos dos cursos são padronizados nacionalmente, partiu-se de máquinas e equipamentos que são imprescindíveis para execução das próprias aulas práticas ao longo do processo formativo.

As duas categorias escolhidas para detalhamento foram: a) infraestrutura necessária para a montagem de provas por escola, contemplando as subseções de equipamentos em geral, instrumentos, material de escritório e ferramentas; e b) insumos necessários para execução de prova por aluno, considerando as subseções de material de consumo e Equipamentos de Proteção Individual (EPI). A principal diferença entre elas está em que o foco da primeira está no posto de trabalho, ou seja, os recursos poderão ser compartilhados mais de uma vez e em momentos de aplicação distintos, enquanto da segunda são aqueles que serão utilizados uma única vez, exclusivamente, por cada estudante que vier realizar a prova. No Anexo B é possível verificar a lista de infraestrutura para o curso técnico em mecânica.

3.5.3 Manual do Avaliador

O manual do avaliador é um documento norteador que auxilia na preparação, aplicação e correção das provas práticas. É desenvolvido pelos próprios elaboradores dos instrumentos de avaliação e contém orientações sobre a organização do ambiente físico, preparação da avaliação, procedimentos a serem observados durante a execução da prova e observações quanto aos critérios de correção das evidências. Além disso, traz o *layout* e como as máquinas, equipamentos e demais insumos devem estar dispostos no ambiente.

A proposição de um manual surge da necessidade de contextualizar, numa visão holística, os principais requisitos para realização efetiva do teste conforme idealização dos autores. Apresentando linguagem direta, este material direciona as ações e cuidados em que os avaliadores devem praticar antes, durante e posteriormente ao processo de aplicação da prova. No Anexo C é possível consultar um recorte deste documento.

3.5.4 Lista de verificação

Este é o instrumento que elenca os critérios de avaliação a serem observados no desempenho do estudante, configurados no formato de evidências. É de uso exclusivo do avaliador e guiam a investigação quanto ao desenvolvimento das capacidades descritas na matriz de referência. São elaboradas com base na situação-problema a ser solucionada, analisadas pelo prisma do domínio psicomotor, utilizando-se como objeto de estudo os produtos desenvolvidos pelo discente ou o modo como executam as atividades.

A avaliação ocorre simultaneamente ao processo de aplicação quando relacionadas as evidências de processo e posteriormente quando associadas ao resultado do trabalho esperado, isto é, evidências de produto. São julgadas de modo dicotômico – “Sim” ou “Não” – e carregam o significado de acerto ou erro, se conseguiu realizar uma ação ou não, se seguiu um determinado procedimento ou não. No Anexo D é possível observar como ocorre a diagramação da lista de verificação.

Cada evidência arrolada corresponde exclusivamente a uma única capacidade da matriz de referência. Os elaboradores são também convidados a correlacioná-las com o perfil profissional do curso para melhor interação com as atividades profissionais requeridas pelo mundo do trabalho. Esta estratégia objetiva aumentar a precisão na interpretação do resultado, pois quando o estudante não consegue solucionar um

determinado critério sabe-se com clareza qual capacidade não foi mobilizada de modo adequado. Logo, para os “não” atribuídos ao desempenho do aluno, deve-se proceder com uma justificativa, no intuito de fornecer informações para tratamento e retroalimentação do processo de ensino-aprendizagem.

3.6 PRÉ-TESTE

Após a elaboração dos instrumentos de avaliação foi organizado um pré-teste para validação das estratégias concebidas. Este ocorreu em uma escola selecionada de acordo com critérios de disponibilidade, em ambiente real (laboratório ou oficina) conforme as características demandadas pela situação-problema. Foram convidados cinco especialistas técnicos, um representante de cada região geográfica do Brasil, com o mesmo perfil requerido para ser um avaliador desta área. Um deles era o próprio autor dos instrumentos, cuja função era observar a reação dos demais colegas quanto às especificações fornecidas.

Com duração de uma semana, este momento foi seccionado da seguinte maneira: no primeiro dia foi realizada a apropriação dos outros quatro docentes quanto aos instrumentos de avaliação, sua leitura integral e marcação de pontos de dúvidas; no segundo dia aplicou-se a prova com três estudantes que já haviam concluído o curso e, portanto, não seriam alvo de futuras avaliações; e do terceiro ao quinto dia foram realizados ajustes conforme os dados observados no comportamento dos egressos.

As principais características que buscava-se investigar neste processo eram: adequação do tempo de execução das atividades demandadas (compatibilidade com o período de três horas); compreensão da situação-problema quanto ao entendimento do que deveria ser solucionado e quais as entregas esperadas (análise semântica); se havia todas as informações necessárias para sua resolução; e se as atividades propostas permitiam observar o comportamento do estudante conforme as evidências descritas. Além disso, após a aplicação era realizada uma entrevista quanto às percepções do teste, como grau de dificuldade em cada tarefa, pontos de dúvidas durante o processo e se havia todas as informações necessárias.

Os docentes foram convidados a praticarem a lista de verificação, simulando a avaliação com base no resultado dos três estudantes pré-testados, de modo individual, sem consulta ou conversa para consenso. Após cada um preencher seu formulário, as repostas foram comparadas e as divergências discutidas para identificar se haviam problemas quanto ao entendimento da redação, se estavam calibradas adequadamente e,

principalmente, se existiam vieses de subjetividade. Após todos os apontamentos, foram revisitados os documentos com o objetivo de aperfeiçoar e garantir que os problemas identificados fossem sanados. Desse modo, todos os itens (evidências) foram mantidos na versão final da lista de verificação, com pequenos ajustes de redação visando ampliar sua compreensão.

Válido ressaltar também que os especialistas técnicos envolvidos nas etapas de elaboração e testagem dos instrumentos de avaliação devem manter confidenciais todas as informações obtidas ou geradas durante a condução das atividades, sendo obrigatória a assinatura de um termo comprometendo-se a com o sigilo, incluindo cláusulas relativas à confidencialidade, imparcialidade e conflito de interesses.

3.7 APLICAÇÃO DAS PROVAS

Este processo dividiu-se em três etapas: planejamento, execução e registro, conforme as características destacadas nas seções subsequentes. Compete esclarecer que não houve necessidade de solicitar aprovação destes procedimentos junto ao Comitê de Ética da UFSC, dado que a coleta de dados era de responsabilidade da instituição de educação profissional (SENAI), restringindo o papel do pesquisador à utilização dos resultados para compreensão, análise e proposição de uma escala pedagógica interpretativa. Ver autorização da pesquisa no Anexo A.

Também é importante aclarar que a participação dos estudantes foi voluntária e amostral, ou seja, eles poderiam recusar-se a efetuar o teste prático caso fossem sorteados, já que esse processo avaliativo não gerava conceitos ou notas para serem incluídas em sua formação, uma vez que o objetivo da prova é verificar a qualidade da oferta dos cursos.

3.7.1 Planejamento

Inicialmente foi necessário observar os aspectos relacionados à estimativa da amostra para cada escola, homologação dos estudantes (confirmação dos indivíduos sorteados), seleção e capacitação dos avaliadores e familiarização com os instrumentos de avaliação e infraestrutura.

Quanto à definição do tamanho da amostra necessária em cada unidade operacional, utilizou-se como parâmetro o desvio padrão associado a realização da avaliação cognitiva dos estudantes do ano anterior (2016). Para o curso técnico em mecânica, considerando a escala de proficiência (500, 100), o valor de referência foi 100,66. Além disso,

considerou-se a equação do cálculo amostral apresentada no item 3.3 e arbitrou-se como erro aceitável 30 pontos, pois era o menor quantitativo possível frente as condições logísticas e operacionais.

De acordo com a amostra estabelecida, foram randomicamente sorteados os estudantes em número equivalente dentre os matriculados e aptos para realizar a prova, bem como ordenados discentes reserva em caso de indisponibilidade. Este processo foi gerenciado eletronicamente via Sistema de Avaliação Prática (SAP) e cabia aos coordenadores das escolas homologá-los, isto é, confirmar a participação dos alunos selecionados. Admitiam-se, entretanto, algumas possibilidades para não validação como não participação da etapa anterior (avaliação objetiva), evasão/abandono do curso ou por motivos de atestado médico.

Também precisava-se estabelecer quem seriam os avaliadores e aplicadores das provas. Este papel deveria ser conduzido por um profissional com qualificações pessoais/técnicas na área da metalmecânica e competência em avaliar estudantes, observando alguns aspectos essenciais:

- possuir formação ou experiência no curso técnico avaliado;
- apresentar domínio técnico quanto às capacidades formativas previstas na matriz de referência;
- comunicar-se adequadamente;
- demonstrar postura isenta e conduta desprovida de quaisquer interesses, para que possam fazer julgamentos imparciais dos desempenhos observados;
- dominar a utilização das ferramentas informatizadas para a geração das provas e tabulação dos resultados.

Essas pessoas eram capacitadas virtualmente pela plataforma SAP, cujo treinamento foi segmentado em módulos que versavam sobre as temáticas: introdução ao Sistema de Avaliação da Educação Profissional (SAEP), matrizes de referências da avaliação, situações de aprendizagem, instrumentos de avaliação e ambientação dos avaliadores. Além disso, com quinze dias de antecedência à aplicação, foram disponibilizados os instrumentos de avaliação para familiarização. O objetivo era apropriar-se corretamente dos desafios propostos e dos resultados esperados, contando com webconferências com o elaborador da prova para sanar dúvidas e orientar em apontamentos não contemplados anteriormente.

No quesito infraestrutura, o avaliador era responsável por preparar com antecedência o ambiente em que o estudante realizaria o teste, conferindo o atendimento de todas as determinações previstas no manual,

tais como equipamentos, ferramentas, insumos, montagem da atividade e documentações técnicas/normativas para consulta dos avaliados.

3.7.2 Execução

A aplicação das provas práticas dos 1.498 estudantes ocorreu no período de 25 de setembro a 06 de outubro de 2017. Cada escola organizou, conforme a capacidade de infraestrutura e disponibilidade do avaliador, a melhor data e a quantidade de estudantes que seriam simultaneamente testados.

Antes mesmo de iniciar a avaliação, foram oportunizados 30 minutos para ambientação e orientação, englobando informações gerais do teste, familiarização com a infraestrutura/layout do ambiente, bem como a leitura do caderno de prova. Eram permitidos questionamentos ou esclarecimentos sobre qualquer um dos itens, com exceção aos requisitos técnicos de competência prevista no perfil profissional ou alguma informação que o auxiliasse nas respostas. Ainda, conferia-se a identidade e autoria (mediante documento oficial com foto) e realizava a coleta da assinatura.

Ao esgotar esse prazo, os discentes possuíam até três horas para executar todas as tarefas previstas no caderno de prova, de modo autônomo, sem utilização de equipamentos eletrônicos ou outros meios de consulta não disponibilizados. Da mesma forma, a comunicação se restringia ao esclarecimento de dúvidas exclusivamente direcionadas ao professor. Além do mais, não eram permitidas interferências no modo como o estudante desenvolvia as atividades, excetuando os casos onde haviam exposição a riscos de acidentes oriundos da negligência no uso de equipamentos/ferramentas ou pela falta de proteção/segurança em sua conduta. Situações como essas e o comportamento dos sujeitos eram verificados ao longo da aplicação, enquanto as evidências de produtos somente eram examinadas com a conclusão do prazo. No manual há anexos e gabaritos que permitiram maior precisão e clareza na atribuição do “sim” ou “não” quanto aos critérios avaliativos.

Ao final desta etapa o caderno de prova e demais itens que compõem os instrumentos de avaliação deviam ser recolhidos, incluindo os produtos e/ou entregas geradas pelos alunos. Também se instruiu sobre a guarda em meio físico por, no mínimo seis meses, contados do final do período de aplicação das avaliações práticas.

3.7.3 Registros da aplicação

O principal artifício que possibilitou o registro foi a lista de verificação, onde cada evidência deveria ser julgada quanto ao atendimento ou não (avaliação dicotômica), baseada exclusivamente no desempenho do estudante. Primeiramente efetuou-se o apontamento manual, pois uma parcela das análises ocorrera concomitantemente a execução, para posterior lançamento eletrônico na plataforma SAP. Este documento deveria ser preenchido fisicamente e assinado pelo docente, anexando também uma versão digitalizada no ambiente virtual.

Para cada “não” atribuído havia a necessidade de descrever uma breve justificativa que descrevesse o motivo pelo qual o aluno não conseguiu evidenciar aquela determinada capacidade. Sua redação precisava ser objetiva, clara e específica, para que posteriormente auxiliasse na interpretação dos resultados e possibilitasse identificar as dificuldades/erros mais comuns ocorridos.

Uma vez que os registros eram migrados para a plataforma digital, o sistema não permitia finalizar um lançamento sem que todos os campos estivessem preenchidos, alertando os docentes caso alguma lacuna estivesse em branco. Por esse motivo todos os estudantes estavam com dados completos, não tendo sido descartada nenhuma informação ou sujeito para análise dos resultados. É possível verificar na Figura 11 a função de registro da aplicação na plataforma SAP.

Figura 11 - Captura de tela da Plataforma SAP (exemplo fictício)

SAP 2018
Sistema de Avaliação Prática

Site de Testes e Homologação

Home Central de Preparação Fórum Homologação Agendamento Processamento de Resultados Escotas Dados Pessoais Usuários

Confirmação do Coordenador

Informações Gerais

Certificação: TÉCNICO EM ELETROTÉCNICA Candidato: JOSÉ DANIEL DE ALMEIDA SANTOS CPF: 392.919-910-13
Exame realizado em: 31/05/2018 Hora: 23:59 Local: Laboratório 1 - Eletrológica
Avaliador: MARCOS ANDRÉ DOS SANTOS FERREIRA

Painel de Evidências Observáveis

Observação: As evidências observáveis em negrito e sublinhado são críticas.

Elemento de Competência	Padrão de Desempenho	Evidência observável	Sim	Não	Justificativa
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Elaborar a montagem de equipamentos, componentes e circuitos de força, iluminação, comunicação, controle, alarmamento, automação industrial etc.	Executou fielmente as conexões no termo de montagem.	<input type="checkbox"/>	<input type="checkbox"/>	Descrição de Justificativa
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Elaborar a montagem de equipamentos, componentes e circuitos de força, iluminação, comunicação, controle, alarmamento, automação industrial etc.	Utilizou terminais em todas as conexões de acordo com a seção dos condutores.	<input type="checkbox"/>	<input type="checkbox"/>	Descrição de Justificativa
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Elaborar a montagem de equipamentos, componentes e circuitos de força, iluminação, comunicação, controle, alarmamento, automação industrial etc.	Executou a aperto dos bornes e conexões de instalação.	<input type="checkbox"/>	<input type="checkbox"/>	Descrição de Justificativa
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Elaborar instalações de sistemas elétricos industriais.	Descartou os resíduos adequadamente.	<input type="checkbox"/>	<input type="checkbox"/>	Não descartou os resíduos no tempo
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Caracterizar destinação correta dos resíduos.	Instalou disjuntor ou fusível para a proteção do circuito.	<input type="checkbox"/>	<input type="checkbox"/>	Descrição de Justificativa
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Instalar o dispositivo de proteção de sistema elétrico industrial.	Realizou o comissionamento de sistema elétrico.	<input type="checkbox"/>	<input type="checkbox"/>	Descrição de Justificativa
<u>Instalar sistemas elétricos prediais, industriais e de potência (SEP), cumprindo regulamentações vigentes, parâmetros de eficiência energética, normas técnicas, de qualidade, de segurança e saúde e, ainda, ambientais.</u>	Elaborar instalações de sistemas elétricos industriais.	Realizou teste de continuidade de aterramento (PTE) nas instalações e nos metais dos equipamentos.	<input type="checkbox"/>	<input type="checkbox"/>	Descrição de Justificativa

Fonte: SENAI (2018)

3.8 TRATAMENTO DOS DADOS

Esta etapa contou com a análise descritiva da amostra, análise do instrumento, calibração das evidências e estimação das proficiências. Inicialmente buscou-se compreender melhor o perfil do grupo de estudantes avaliados, focando-se nas variáveis localidade, sexo e idade, devido a disponibilidade destas informações na base de dados fornecida pelo SENAI. Utilizou-se o programa Microsoft Excel 16 para resultar a tabulação e geração do gráfico.

Na continuação, trabalhou-se na análise e mensuração da fidedignidade do instrumento, valendo-se do alfa de Cronbach pelo apontamento da literatura para um método de verificação da consistência interna do teste, isto é, se a medida desenvolvida foi confiável quanto a semelhança das evidências direcionadas a um determinado traço latente. Nunally (1978) e Martins e Theóphilo (2009) apontam que coeficiente α superior a 0,70 indica confiabilidade apropriada. Outra medida utilizada como referência nessa vertente é a estabelecida por Hambleton (2004) sobre a CIT, afirmando que o teste apresenta precisão psicométrica quando desempenha nível de informação superior a 10 pontos em algum intervalo de proficiência (θ).

Além do descrito no parágrafo anterior, propôs-se a avaliar a dimensionalidade do teste aplicando duas outras estratégias: análise de componentes principais – baseada no pareamento de itens, adotando-se as correlações tetracóricas pela natureza dicotômica das respostas e análise paralela pela observância no gráfico *scree plot*; e a análise fatorial, conforme valores obtidos pelas cargas fatorais comparativas aos modelos unidimensional e bidimensional. Estes procedimentos foram executados com auxílio do pacote *mirt* e *psych* do R (ver Apêndice A). Cabe destacar que a base de dados precisou ser codificada para realização destes e dos próximos estudos, de modo dicotômico e convertida para a métrica binária, sendo atribuído 1 (um) quando o estudante conseguia atender o critério da evidência (“sim”) e 0 (zero) quando não atendesse (“não”), processada em planilha eletrônica.

Para prosseguimento das etapas subsequentes, era necessário definir qual o modelo de TRI seria empregado. Considerando que a pesquisa possui como natureza a avaliação de uma única população (os estudantes) e um único traço latente (domínio psicomotor), fixou-se como hipótese a máxima de quanto maior a habilidade maior seria a probabilidade de atendimento das evidências. Ainda nesse contexto, o parâmetro “c”, que representa acerto casual, não foi contemplado por não se admitir que um sujeito com baixa proficiência conseguisse realizar

entregas reais de produtos ou processos inerentes as atividades profissionais, uma vez que as evidências somente são alcançáveis quando partem da efetividade do desenvolvimento da capacidade. Logo, apenas poderiam ser adotados modelos que envolvessem dois parâmetros (discriminação e dificuldade). Cruzadas essas informações, conforme Figura 8, a seleção da medida mais apropriada foi o Modelo Logístico de Dois Parâmetros (ML2P). Portanto, os elementos da equação (1), agora no contexto da avaliação do domínio psicomotor, se traduzem da seguinte forma:

- U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j atende ao critério previsto na evidência i , ou 0 quando o indivíduo j não atende ao critério previsto na evidência i .
- θ_j representa a proficiência do domínio psicomotor do j -ésimo indivíduo.
- $P(U_{ij} = 1 | \theta_j)$ é a probabilidade de um indivíduo j com proficiência do domínio psicomotor θ_j atender a evidência i e é chamada de FRI.
- a_i é o parâmetro de discriminação (ou de inclinação) da evidência i , com valor proporcional à inclinação da CCI no ponto b_i .
- b_i é o parâmetro de dificuldade (ou de posição) da evidência i , medido na mesma escala da proficiência do domínio psicomotor.

O software empregado para produção dos resultados foi o BILOG MG®, produzido pela *Scientific Software International*. Este é um programa computacional desenvolvido especificamente para cálculos envolvendo modelos logísticos da TRI. Para processar os cálculos estabelece-se uma sintaxe característica demandando códigos específicos para o número de itens, alternativas de resposta e método de estimação, dentre outros (consultar o Apêndice A).

O aplicativo executa as análises em três etapas:

- fase 1: verificação da entrada de dados e análise clássica das evidências;
- fase 2: calibração das evidências; e
- fase 3: estimação das proficiências.

A primeira fase caracterizou-se pela leitura das informações, quando foram inseridos a identificação dos indivíduos (por máscaras) com suas respectivas respostas já corrigidas, isto é, sistematizados de modo binário (ver Anexo F). Nesta etapa são calculadas algumas estatísticas descritivas a respeito de cada item, como número total de indivíduos avaliados, quantidade de acertos por evidência, além das correlações bisserial. Nas palavras de Andrade, Tavares e Valle (2000), a relevância dessa etapa não está somente em verificar se a interpretação dos dados, por parte do programa computacional, é coerente, mas também em servir como ponto de partida para os processos de estimação das fases subsequentes.

Ainda nessa temática, Soares (2005) aponta este mesmo coeficiente como uma medida estatística onde se compara o resultado de uma evidência em particular da prova com o resultado do teste como um todo, devendo ser interpretada como um indicador de discriminação do item em relação ao escore bruto total. Santos (2008) reflete ainda que esta será considerada discriminativa quando os sujeitos que vão bem no teste acertam-na e os indivíduos com pouco desempenho erram-na. Logo, quando maior seu valor, maior será a capacidade em distinguir com qualidade os estudantes proficientes dos não proficientes. Portanto, correlações baixas ou negativas indicam problemas de entendimento, seja com a construção da própria evidência ou com o juízo do respondente (ANDRADE, TAVARES e VALLE, 2000). O INEP (2012) traz uma referência para sua classificação, caracterizando valores menores que 0,19 como “fraco”, entre 0,20 a 0,29 como “médio”, 0,30 a 0,39 como “bom” e acima de 0,40 como “muito bom”.

Outra estatística a ser considerada nesse ciclo refere-se ao percentual de acerto em cada uma das evidências avaliadas. Conforme propõe Conde (2001), a interpretação para a dificuldade com base na análise TCT pode seguir o seguinte preceito: é denominada “fácil” se seu valor for maior que 0,70; “média” se o índice oscilar entre 0,30 e 0,70; e qualificada como “difícil” se o número estiver abaixo de 0,30.

Já na segunda fase realizou-se a calibração das evidências, isto é, foram estimados os parâmetros de discriminação e dificuldade (“a” e “b”), com os seus respectivos erros-padrão. O procedimento escolhido para esta finalidade foi o método da máxima verossimilhança marginal, sendo que o mecanismo empregado para resolução das equações foi o algoritmo EM e o método “*Scoring*” de Fisher. De acordo com Andrade, Tavares e Valle (2000), isto somente é possível por meio da utilização de distribuições de probabilidades para o traço latente. O BILOG-MG parte do pressuposto que os estudantes representavam uma amostra aleatória da

totalidade de proficiências que possam ser alcançadas. Tal situação é apresentada no formato de distribuição discreta por meio de pontos de quadratura.

Para análise da qualidade dos parâmetros utilizou-se os critérios apontados por Baker (2001) para classificação dos índices de discriminação, admitindo que os valores de “a” possam ser organizados na seguinte escala (0, 1): “nenhuma” para valores negativos ou zerados; “muito baixa” entre 0,01 e 0,34; “baixa” entre 0,35 e 0,64; “moderada” entre 0,65 e 1,34; “alta” entre 1,35 e 1,69; e “muito alta” quando maior que 1,70. No entendimento desta pesquisa, compreendeu-se como aceitável aquelas diagnosticadas de modo equivalente ou superior a categoria “moderada”. Já em relação ao indicador de dificuldade “b”, Vendramini e Dias (2005) sugerem que seja observado o intervalo de -2,95 a 2,95 como adequado, entretanto significâncias maiores ou menores não indicam necessariamente problemas, pois quanto mais variabilidade houver na escala melhor para compreensão da habilidade.

Ainda nesta etapa gerou-se a análise gráfica, contemplando a Curva Característica do Item (CCI), Curva de Informação do Item (CII) e a Curva de Informação do Teste (CIT). A primeira permitiu assimilar principalmente o comportamento da evidência quanto sua inflexão. Já a segunda aponta o nível de informação e contribuição de cada item para explicação do fenômeno. Por fim, a terceira sugere a faixa de desempenho onde o teste apresenta maior relevância.

A terceira e última fase relacionou-se com o procedimento de estimação do traço latente de cada estudante avaliado, representado pelo parâmetro θ e calculado com base na fase anterior. Para esta pesquisa, a estimação da habilidade do respondente era alusiva aos dois parâmetros do modelo logístico: “a” – capacidade de a evidência distinguir a proficiência do domínio psicomotor entre os respondentes de alta e baixa habilidade e “b” – o nível de proficiência do domínio psicomotor necessário para que a evidência fosse marcada como “sim” pelo avaliador; e

Como procedimento técnico empregou-se o método EAP. Nessa abordagem as habilidades são quantificadas aplicando pontos de quadratura para aproximar a distribuição a priori com o desempenho de cada estudante. Utilizou-se 20 pontos, sendo que o *software* acumulou as densidades de todos os sujeitos, normalizou-as e gerou probabilidades em cada ponto, demonstrando que o número definido era adequado para a análise deste conjunto de dados.

Outros comandos foram inseridos na sintaxe para estimação das proficiências, sendo: NQPT=20, CYCles=50, Newton=0, CRIt=0.001. O

primeiro descreve a quantidade de pontos de quadratura, descrito no parágrafo anterior. O segundo código se refere ao número de ciclos para estabilização das análises, utilizando-se 50 como sugestão inicial e avaliando se seu resultado era suficiente para atendimento desta premissa. O terceiro era uma instrução para não utilização do algoritmo *Newton-Raphson*, uma vez que o método escolhido foi o EAP (expresso na sintaxe por `METHOD=2`). Já o último é em relação ao critério de precisão, adotando três números após a vírgula para aumentar a segurança e não enviesar os processos de estimação.

3.9 CONSTRUÇÃO, INTERPRETAÇÃO E APLICAÇÃO DA ESCALA

Para elaboração da escala de proficiência do domínio psicomotor foram utilizados dois procedimentos principais: identificação de evidências âncora fundamentadas por análises estatísticas e a apresentação das evidências calibradas para especialistas com domínio técnico na área avaliada, doravante denominados de critério estatístico e critério técnico, respectivamente.

No primeiro, foram revisitadas as métricas propostas por Andrade, Tavares e Valle (2000) adaptando-as para o conjunto de dados, já que este não considerou o parâmetro de acerto ao acaso e que a ancoragem seria realizada a cada 0,5 desvios-padrão. Por definição arbitrária, estabeleceu-se o seguinte modo para posicionamento nos níveis âncoras: $P(U = 1|\theta = Z) \geq 0,60$ e $P(U = 1|\theta = Y) < 0,50$. Na segunda etapa, os especialistas foram convidados a alocar as evidências entre duas opções: nível em que a probabilidade de θ é maior ou igual a 0,60 e o ponto imediatamente anterior, escolhendo a posição que melhor harmonizasse com as demais já pré-fixadas (resultado consolidado na Figura 12).

Figura 12 - Posicionamento das evidências com especialistas



Fonte: elaborado pelo autor

Outro ponto a ser observado diz respeito à conversão da escala de (0, 1) para (100, 10), ou seja, média 100 e desvio padrão 10. Esta escolha foi conduzida de modo discricionário, uma vez que seria acompanhada de uma interpretação pedagógica para conexão entre os valores estimados e o conteúdo existente nas evidências. Para transformação dos valores utilizou-se como base as seguintes equações:

$$\theta^* = 10 \times \theta + 100 \quad (3)$$

$$P(U_i = 1|\theta) = P(U_i = 1|\theta^*) \quad (4)$$

Onde o θ^* representa o valor da habilidade recalculado para a medida (100, 10).

Definidos os valores numéricos a serem representados em cada ponto da escala, sucedeu-se com a análise pedagógica de cada nível, considerando espaçamento de 0,5 desvios-padrão para cada uma das lateralidades da escala. Esta análise foi alicerçada pelos descritores das próprias evidências, ponderadas pela análise técnica dos especialistas. Na sequência, estabeleceu-se as divisões dos níveis de desempenho conforme a necessidade apontada pelos docentes para interpretação, convidados a definir de três a cinco níveis que fossem diferentes da classificação adotada pela instituição na avaliação do domínio cognitivo (“abaixo do básico”, “básico”, “adequado” e “avançado”).

A descrição pedagógica dos níveis contemplou dois momentos. Primeiramente uma abordagem sumária, sintetizando as principais capacidades mobilizadas pelos estudantes em cada faixa de desempenho. E depois, numa perspectiva mais pormenorizada, comparando quais capacidades são mobilizadas de acordo com a matriz de referência. Contudo, devido à natureza cognitiva da matriz de referência utilizada (constructo), os especialistas foram convidados a analisar cada uma das evidências e reescrevê-las de acordo com o domínio psicomotor mobilizado, promovendo nova classificação e agrupamento distintos do previsto originalmente.

Um desafio nesse sentido é a ausência de suporte teórico que oriente como uma capacidade do nível psicomotor deva ser escrita no contexto da educação profissional. Exemplo disso é o trabalho de Harrow (1983), pois traz exemplo de taxonomia nessa vertente, porém para aplicações na educação física e no desenvolvimento da motricidade da criança, não se enquadrando como adequadas frente à complexidade de movimentos motores que um trabalhador do segmento da mecânica

executa. A proposta estabelecida por Dave (1970) e Simpson (1972) foi priorizada nesta pesquisa por retratar melhor o processo evolutivo da aprendizagem desta dimensão, ainda que também não apresente indicação de formulação das capacidades. Por esse motivo, optou-se por utilizar verbos que estivessem relacionados ao “fazer” e que demonstrassem que tipo de nível de domínio psicomotor estava sendo mobiliado.

Outrossim, para estabelecer o detalhamento da segunda etapa da descrição pedagógica utilizou-se a proposta de Mager (1979) para formulação de um objetivo educacional específico, que contempla: descrição do comportamento final ou a conduta observável do aprendiz; definição do critério de qualidade ou nível de desempenho satisfatório; e delimitação da condição na qual se espera que o comportamento ocorra. Trazendo para o contexto deste trabalho, adaptou-se essa estrutura para a seguinte formulação: <capacidade psicomotora> **conforme** <critério de avaliação (evidência)>, **considerando** <a condição>. Este formato de redação permeou a tradução de cada um dos níveis conforme posicionamento realizado na construção da escala.

Por fim, converteu-se a proficiência de cada estudante para a nova métrica (100, 10) e classificou-os nos níveis de desempenho definidos, aplicado a distintos contextos: visão macro, análise por unidade federativa, por sexo e idade. A intenção estava em analisar os diferentes tipos de comportamento da amostra considerando as variáveis associadas aos estudantes disponíveis na base de dados.

Esclarecidos todos os procedimentos metodológicos, parte-se para a investigação dos resultados obtidos ao longo da pesquisa, discutindo cada uma das etapas e confrontando-as com os preceitos basilares identificados no segundo capítulo.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise e discussão dos resultados desta pesquisa foi estruturada em sete etapas. Inicialmente apresentou-se algumas estatísticas descritivas, expondo as principais características da população e da amostra avaliada conforme as variáveis disponíveis para estudo. Na sequência, procedeu-se com a calibração dos parâmetros das evidências, aplicando a TRI e identificando os valores correspondentes dos parâmetros de discriminação e dificuldade. A partir disso, possibilitou-se a estimação da proficiência de cada indivíduo para o domínio psicomotor, considerando os dados obtidos na fase anterior. Na quarta etapa realizou-se a verificação da validade e fidedignidade do instrumento com base na fundamentação teórica já elencada anteriormente. Posteriormente, posicionou-se as evidências na escala de proficiência, definindo os cortes para identificação dos níveis de desempenho, assim como sua interpretação pedagógica para que docentes e gestores escolares pudessem utilizá-la na melhoria dos processos educacionais. Por fim, fez-se a aplicação da escala criada para a amostra, observando o comportamento dos diferentes grupos quanto aos graus de traço latente mensurado. Os subcapítulos a seguir detalham cada passo destes com toda sua amplitude.

4.1 ANÁLISE ESTATÍSTICA DESCRITIVA

O propósito desta etapa está em analisar algumas estatísticas descritivas que permitam caracterizar o público avaliado. Assim, a população inicial totalizava 5.856 estudantes da educação profissional (Curso Técnico em Mecânica), porém ponderando as características logísticas e operacionais das provas práticas, optou-se por trabalhar de forma amostral aplicando a equação (2), resultando 1.498 discentes selecionados. Este quantitativo representa 25,6% do grupo original e aceitou-se como erro amostral 30 pontos na métrica da escala de proficiência da prova cognitiva (500, 100).

As variáveis selecionadas para entender o comportamento desta amostra e que estavam disponíveis para utilização do pesquisador foram: Unidade Federativa (UF) de origem, sexo e idade. Quanto à classificação das variáveis em relação ao nível de mensuração temos:

- UF: considerada qualitativa nominal, pois admitia-se apenas o estado de origem do estudante, podendo assumir 16 valores sem relação ordinal, sendo: AM (Amazonas), BA (Bahia), CE

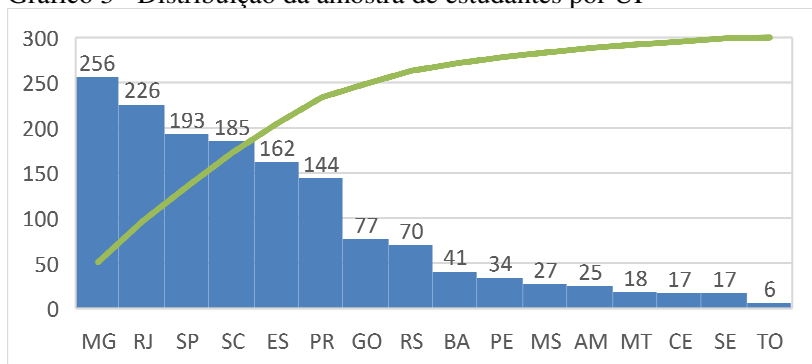
(Ceará), ES (Espírito Santo), GO (Goiás), MG (Minas Gerais), MS (Mato Grosso do Sul), MT (Mato Grosso), PE (Pernambuco), PR (Paraná), RJ (Rio de Janeiro), RS (Rio Grande do Sul), SC (Santa Catarina), SE (Sergipe), SP (São Paulo) e TO (Tocantins).

- Sexo: identificada de mesma forma como qualitativa nominal, porque expressa o sexo biológico do respondente, aceitando apenas dois valores: masculino ou feminino.
- Idade: destacada como quantitativa discreta, optando-se pela apresentação dos dados em anos completos de vida, variando entre 18 e 63 anos.

Importante salientar que não houve a necessidade de realizar higienização da base de dados em razão das informações já estarem padronizadas pelo SAP. Isto porque, somente era possível finalizar o registro de um discente avaliado quando todos os campos estavam preenchidos conforme as categorias mencionadas, não havendo, portanto, lacunas ou incompletudes.

Quanto a localidade (UF), os alunos estavam distribuídos conforme demonstrado no Gráfico 3. O estado com maior representatividade é Minas Gerais (17,1%) enquanto o com menor é Tocantins (0,4%). Aplicando o critério Pareto (sinalizado pela linha verde), os estados que totalizam aproximadamente 80% da distribuição são, sequencialmente, MG, RJ, SP, SC, ES e PR (77,8% de frequência acumulada).

Gráfico 3 - Distribuição da amostra de estudantes por UF



Fonte: elaborado pelo autor

Em relação ao agrupamento por sexo, 1.367 (91,3%) dos estudantes avaliados eram homens, enquanto apenas 131 (8,7%) eram mulheres. Cruzando esta variável por UF, os estados CE e RS possuem 100% dos alunos pertencentes ao sexo masculino, enquanto TO é o que apresenta maior proporção de mulheres (50%), embora seu número absoluto seja baixo. Nessa perspectiva, MG é o mais representativo do público masculino (232) e ES do feminino (30). Mais detalhes na Tabela a seguir.

Tabela 1 - Distribuição da amostra de estudantes por UF e Sexo

UF	Masculino	Feminino	Total	% M	% F
AM	24	1	25	96,0%	4,0%
BA	36	5	41	87,8%	12,2%
CE	17	0	17	100,0%	0,0%
ES	132	30	162	81,5%	18,5%
GO	70	7	77	90,9%	9,1%
MG	232	24	256	90,6%	9,4%
MS	26	1	27	96,3%	3,7%
MT	17	1	18	94,4%	5,6%
PE	28	6	34	82,4%	17,6%
PR	132	12	144	91,7%	8,3%
RJ	220	6	226	97,3%	2,7%
RS	70	0	70	100,0%	0,0%
SC	177	8	185	95,7%	4,3%
SE	15	2	17	88,2%	11,8%
SP	168	25	193	87,0%	13,0%
TO	3	3	6	50,0%	50,0%
Total	1367	131	1498	91,3%	8,7%

Fonte: elaborado pelo autor

No que tange à idade, utilizou-se o critério de Montgomery e Runger (2003), pois “a distribuição de frequências em classes é apropriada para apresentar dados quantitativos contínuos ou discretos com um número elevado de possíveis valores” (Medronho *et al.*, 2003, p. 231). Observando os dados apresentados na Tabela 2, percebe-se que a

maioria dos estudantes possuem idade entre 18 e 25 anos (70,9%) e a minoria está entre aqueles com idade entre 58 e 63 anos (0,1%). Também se encontrou uma média de 24,4 anos e valor mais frequente (moda) de 20 anos. Na divisão por quartil, o primeiro resultou em 20, enquanto o terceiro foi de 27, além da mediana ser 22 anos. Quanto à amplitude, o menor valor encontrado foi 18 e o maior era 63 anos, além de apresentar variância de 40,1 e desvio padrão de 6,3.

Tabela 2 - Distribuição da amostra de estudantes por UF e Idade

Classes --	Frequência	%	Frequência Acumulada	% Acumulado
18 -- 26	1062	70,9%	1062	70,9%
26 -- 34	285	19,0%	1347	89,9%
34 -- 42	118	7,9%	1465	97,8%
42 -- 50	26	1,7%	1491	99,5%
50 -- 58	5	0,3%	1496	99,9%
58 -- 63	2	0,1%	1498	100,0%

Fonte: elaborado pelo autor

4.2 ANÁLISE E CALIBRAÇÃO DAS EVIDÊNCIAS

Inicialmente, antes mesmo da calibração das evidências, fez-se necessária uma análise do conjunto de dados pela TCT. Este procedimento visou verificar se havia algum indicativo de incoerência na interpretação computacional do software utilizado e correspondeu a primeira fase de testes do Bilog-MG. Foram descritos, portanto, número total de indivíduos e itens, quantidade de acertos/erros, além das correlações bisseriais.

4.2.1 Análise pela TCT

Preliminarmente é demonstrado na Tabela 3 a proporção de erro e de acerto de cada evidência (Ev.) para os 1.498 respondentes, isto é, o índice de facilidade. Isso significa dizer que, quanto mais os indivíduos acertaram-na, mais fácil ela era, e quanto menos repostas assertivas (Resp. Cor.), maior a dificuldade enfrenta no momento de demonstrar a habilidade esperada, ou seja, mais difícil.

A evidência com maior percentual de acerto identificada foi a de número 20 “utilizou os equipamentos de proteção individual (óculos de

proteção, calçado de segurança e proteção para cabelos”, quando 95,7% dos estudantes demonstraram atendimento satisfatório. Já a de menor valor é a 06 “representou no desenho os furos de centro conforme anexo 1”, sendo que apenas 40,2% obtiveram “sim” na avaliação do docente.

Aplicando o procedimento definido por Conde (2001), 24 das evidências podem ser classificadas como “fáceis”, pois apresentaram proporções de acerto superiores a 0,70. São elas: 1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30 e 32. As demais são apontadas como de dificuldade média, com valores intermediários entre 0,70 e 0,30, representadas pelos números: 4, 6, 7, 17, 18, 19, 29 e 31. Não foram encontradas evidências consideradas difíceis (inferiores a 0,30), demonstrando preliminarmente que o conjunto de dados tendeu para um teste com resolução mais facilitada.

Tabela 3 - Proporções de erros/acertos por evidência avaliada

Ev.	Resp. Cor.	Proporções		Ev.	Resp. Cor.	Proporções	
		Acerto	Erros			Acerto	Erros
E01	1143	0,763	0,237	E17	766	0,511	0,489
E02	1110	0,741	0,259	E18	792	0,529	0,471
E03	1125	0,751	0,249	E19	988	0,660	0,340
E04	864	0,577	0,423	E20	1433	0,957	0,043
E05	1319	0,881	0,119	E21	1387	0,926	0,074
E06	602	0,402	0,598	E22	1094	0,730	0,270
E07	930	0,621	0,379	E23	1116	0,745	0,255
E08	1316	0,879	0,121	E24	1344	0,897	0,103
E09	1330	0,888	0,112	E25	1288	0,860	0,140
E10	1361	0,909	0,091	E26	1395	0,931	0,069
E11	1334	0,891	0,109	E27	1367	0,913	0,087
E12	1291	0,862	0,138	E28	1386	0,925	0,075
E13	1315	0,878	0,122	E29	1001	0,668	0,332
E14	1106	0,738	0,262	E30	1239	0,827	0,173
E15	1105	0,738	0,262	E31	985	0,658	0,342
E16	1409	0,941	0,059	E32	1274	0,850	0,150

Fonte: elaborado pelo autor

Na sequência, se observou as correlações bisseriais (Biss.) resultante da primeira rodada de cálculos adjacentes aos recursos

computacionais viabilizados para condução do estudo, estabelecidos no Quadro 4.

Quadro 4 - Correlação bisserial das evidências avaliadas

Ev.	Biss.	Ev.	Biss.	Ev.	Biss.	Ev.	Biss.
E01	0,634	E09	0,927	E17	0,412	E25	0,816
E02	0,521	E10	0,966	E18	0,403	E26	0,968
E03	0,613	E11	0,928	E19	0,376	E27	0,768
E04	0,545	E12	0,794	E20	0,792	E28	0,882
E05	0,786	E13	0,895	E21	0,855	E29	0,644
E06	0,487	E14	0,556	E22	0,768	E30	0,741
E07	0,448	E15	0,483	E23	0,741	E31	0,683
E08	0,881	E16	0,727	E24	0,816	E32	0,694

Fonte: elaborado pelo autor

As correlações podem ser classificadas de modo geral como “muito bom”, de acordo com a metodologia INEP (2012) e somente a E19 qualificada como “bom”. Logo, não houve nenhuma com valor negativo, podendo inferir previamente que não há problemas de entendimento relativo ao processo de construção dos critérios avaliativos. Ressalta-se ainda que todos os coeficientes bisseriais situaram-se acima de 0,37, demonstrando que de fato apresentam consistência interna e se associam bem ao conjunto.

Com base nessas interpretações, verificou-se que não era necessário excluir nenhuma evidência dos processos subsequentes, visto que os parâmetros encontrados estão dentro dos aceitáveis estabelecidos pela literatura, permitindo assim avançar para a próxima fase.

4.2.2 Calibração pela TRI

Para estimação dos parâmetros desta fase, o modelo que permitiu análise mais adequada foi o logístico de dois parâmetros (ML2P). Nesse ínterim, buscou-se identificar os índices associados a discriminação (“a”) – que identifica quanto os sujeitos de diferentes habilidades de domínio psicomotor diferem em relação a probabilidade de acerto a uma evidência – e a dificuldade (“b”) – relativa à frequência de acertos no computo geral dos estudantes, conforme demonstrado na Tabela 4.

Tabela 4 - Calibração dos parâmetros “a” e “b” das evidências

Ev.	a	E.P (a)	b	E.P (b)	Ev.	a	E.P (a)	b	E.P (b)
E01	1,576	0,100	-1,053	0,057	E17	0,811	0,062	-0,066	0,068
E02	1,186	0,082	-1,121	0,074	E18	0,739	0,061	-0,176	0,075
E03	1,640	0,101	-0,976	0,052	E19	0,580	0,059	-1,227	0,145
E04	1,452	0,084	-0,301	0,043	E20	1,344	0,147	-2,866	0,228
E05	2,304	0,164	-1,461	0,058	E21	1,436	0,131	-2,283	0,148
E06	1,271	0,077	0,406	0,050	E22	1,788	0,106	-0,849	0,045
E07	1,051	0,071	-0,577	0,062	E23	1,647	0,101	-0,947	0,051
E08	4,089	0,299	-1,245	0,033	E24	1,412	0,118	-2,002	0,121
E09	5,009	0,394	-1,256	0,029	E25	1,561	0,115	-1,602	0,084
E10	5,830	0,512	-1,350	0,028	E26	1,681	0,150	-2,146	0,124
E11	4,974	0,394	-1,271	0,030	E27	1,265	0,116	-2,310	0,159
E12	3,009	0,203	-1,242	0,040	E28	1,513	0,135	-2,205	0,136
E13	4,174	0,306	-1,237	0,032	E29	1,376	0,085	-0,689	0,052
E14	1,052	0,077	-1,198	0,086	E30	1,409	0,101	-1,478	0,083
E15	0,858	0,071	-1,386	0,114	E31	1,595	0,093	-0,595	0,044
E16	1,262	0,130	-2,693	0,207	E32	1,232	0,097	-1,774	0,111

Fonte: elaborado pelo autor

Nesta etapa também, o Bilog-MG demonstra a convergência do processo iterativo utilizado (método EM) e a quantidade de ciclos necessários para atingir a convergência (alcançada em 29 ciclos), isto é, sem indicar erros com os dados ou insuficiências na amostragem.

Não há um consenso entre os autores da literatura técnica para avaliação da qualidade dos parâmetros, porém conforme sugestão realizada por Baker (2001), as evidências deste teste, quanto a discriminação, podem ser classificadas de acordo com o exposto no Quadro 5.

De modo geral, quase todas as evidências demonstraram comportamento discriminatório superior a classificação “moderada” (96,9%), entretanto a E19 “executou acabamento superficial do rebaixo D com Ra 3,2” teve desempenho considerado baixo (0,580). Contudo, mesmo realizando novamente a calibração excetuando esta evidência, não houve alterações significativas na estimação dos demais parâmetros, optando-se como decisão mantê-la no conjunto total, mas sem posicionamento posterior na escala de proficiência. Além disso, embora não se tenha referências sobre valores teto, chama-se atenção para E08, E09, E11, E13, e E10 com indicadores próximos ou superiores a 5, sendo esta última “desenhou o corpo central conforme anexo 1” a mais

expressiva de todas (5,830), sendo que errá-las indica que o sujeito possui um nível muito baixo de proficiência.

Quadro 5 - Classificação das evidências quanto ao parâmetro de discriminação

Classificação	Parâmetro	Evidências	Total	%
Muito baixa	0,01 – 0,34	-	-	0,00%
Baixa	0,35 – 0,64	E19	01	3,12%
Moderada	0,65 – 1,34	E02 / E06 / E07 / E14 / E15 / E16 / E17 / E18 / E20 / E27 / E32	11	34,38%
Alta	1,35 – 1,69	E01 / E03 / E04 / E21 / E23 / E24 / E25 / E26 / E28 / E29 / E30 / E31	12	37,50%
Muito alta	1,70 – +∞	E05 / E08 / E09 / E10 / E11 / E12 / E13 / E22	08	25,00%

Fonte: elaborado pelo autor

Cabe refletir ainda que, embora a maioria absoluta estejam dentro dos limites de tolerância aceitáveis, quanto maior for a discriminação, maior o seu poder em distinguir a probabilidade de acerto entre os indivíduos considerados habilidosos dos não habilidosos. Na análise gráfica, por exemplo (ver curva característica do item mais adiante), esse indicador está relacionado diretamente com a inclinação, onde quanto mais íngreme este estiver, mais se consegue diferenciar os indivíduos em relação aos seus traços latentes.

Quanto ao “b”, a única evidência que registrou valor positivo foi a E06 “representou no desenho os furos de centro conforme anexo 1” com 0,406 e, conseqüentemente, caracterizou-se pela mais “difícil” do conjunto, seguidas em ordem decrescente pela 17 “usinou o diâmetro do rebaixo D conforme anexo 3” (-0,066) e 18 “usinou o rebaixo D com o comprimento especificado conforme anexo 3” (-0,176). Já as com menor grau de dificuldade foram, respectivamente, a 20 “utilizou os equipamentos de proteção individual (óculos de proteção, calçado de segurança e proteção para cabelos)” (-2,866), 16 “utilizou instrumento de medição conforme especificação técnica do projeto” (-2,693) e 21 “utilizou os EPIs para realizar a manutenção (óculos de proteção, calçado de segurança e proteção para cabelos)” (-2,283). Observa-se que os

valores mencionados estão baseados numa escala com distribuição normal de métrica (0, 1).

Orientando-se pela sugestão de Vendramini e Dias (2005), para este parâmetro são coerentes valores dentro da faixa de -2,95 a 2,95, sendo que dados fora desse contexto poderiam indicar reflexões quanto ao comportamento de acerto dos estudantes. Percebeu-se que o intervalo de amplitude detectado está entre -2,866 e 0,406 e, portanto, plausível com os pressupostos conceituais. Desse modo, tanto a análise da discriminação quanto a dificuldade não levaram a exclusão de nenhuma evidência do processo estatístico, visto que todas possuem algum nível de informação relevante ao contexto avaliado.

Para sustentar essa tomada de decisão, foram verificados ainda o erro padrão de ambos os parâmetros considerando as dimensões de homogeneidade e magnitude no conjunto. A primeira diz respeito a variabilidade e discrepâncias quando avaliado o todo, onde no caso da discriminação a diferença entre o maior e menor valor é de 0,453 (0,512 – 0,059), enquanto no grau de dificuldade essa distância é ainda menor, com 0,200 (0,228 – 0,028). Dessa forma, pelos valores serem relativamente baixos e aceitáveis, entende-se que em ambas as circunstâncias pode ser considerado homogêneo.

No tocante à segunda, visto que a amostragem segue uma distribuição normal, pode-se afirmar que o verdadeiro valor do parâmetro esteja entre dois erros padrão da estimativa (MONTGOMERY; RUNGER, 2003). Exemplificando, considerando a E01, o verdadeiro valor do parâmetro “a” seria igual a $1,576 \pm 2 \times 0,100$ (entre 1,376 e 1,776) e parâmetro “b” de $-1,053 \pm 2 \times 0,057$ (entre -0,939 e -1,167). Tais valores podem ser interpretados como limites aceitáveis para esse contexto, já que também existe homogeneidade e se permite a manutenção da condição de comparabilidade.

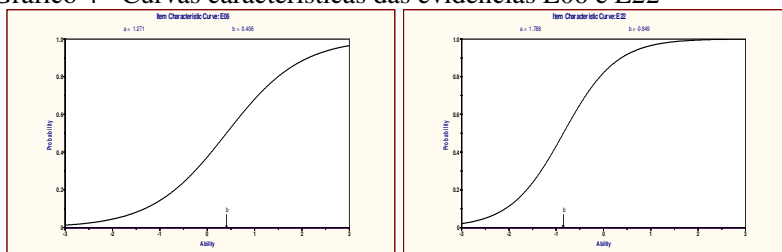
4.2.3 Análise gráfica das evidências

Ainda nesta fase, o programa gera alguns gráficos que ilustram a estimação dos parâmetros e as suas interconexões, como a CCI, CII e CIT. Embora os nomes remetam a “item”, no contexto deste estudo estamos interpretando evidências.

Em relação a CCI, o Gráfico 4 retrata lado a lado o comportamento de duas curvas características selecionadas para fins de comparabilidade e compreensão. O índice “b” representa o ponto de corte no eixo da abscissa em que a probabilidade de resposta assertiva pelo estudante é igual a 0,5. Isso significa dizer que os respondentes com habilidade igual

a 0,406 (na escala 0, 1) possuem 50% de probabilidade de acertar a E06, enquanto na E22 essa mesma proporção será para indivíduos com traço latente de -0,849. Visualmente percebe-se que a segunda curvatura é mais deslocada para a esquerda, até porque o valor de “b” é significativamente menor, demonstrando maior facilidade de resposta pelo conjunto de discente avaliado. Logo, um indivíduo cuja proficiência do domínio psicomotor seja exatamente a média ($\theta = 0$) terá probabilidade de 0,37 para a E06 e 0,82 para a E22, ou seja, haverá maior possibilidade de atendimento do critério avaliativo no segundo caso do que no primeiro, estatisticamente raciocinando.

Gráfico 4 - Curvas características das evidências E06 e E22

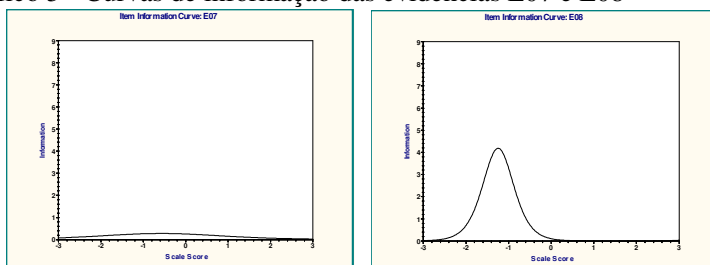


Fonte: elaborado pelo autor

Interpretando o parâmetro “a” com base nas mesmas imagens, percebe-se que este está intrinsecamente correlacionado à derivada da tangente da curva no ponto “b”, ou melhor explicando, onde ela muda de concavidade (também denominado de ponto de inflexão). Por conseguinte, a evidência E22 (1,788) denota uma inclinação superior quando comparada com a E06 (1,271), significando dizer que no primeiro caso há maior propensão de discriminação da proficiência dos estudantes do que no segundo, sendo que cada uma delas possui sua área de maior informação.

Na sequência, verificou-se as CII conforme recorte destacado no Gráfico 5, com o intento de analisar o quanto de informação psicométrica cada evidência abrange no contexto da medida.

Gráfico 5 - Curvas de informação das evidências E07 e E08



Fonte: elaborado pelo autor

Para auxiliar nessa compreensão, utilizou-se as ponderações de Baker (2001) concernente a dois pontos: (1) quanto maior for a discriminação de uma determina evidência, maior será também a informação que ela traz para o θ ; e (2) a informação da será mais característica quando o valor de “b” for igual ao valor do traço latente, pois assim a equação abaixo se satisfaz.

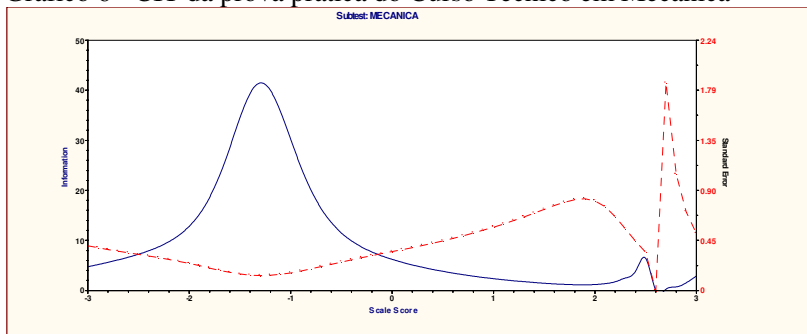
$$\theta - "b" = 0 \quad (5)$$

Ao observar este gráfico, constata-se que a E07 atinge seu ponto máximo na CII quando a habilidade é de -0,577 (mesmo valor do parâmetro “b”), enquanto no segundo exemplo é -1,245. Repara-se também que esta precisão diminui conforme avança-se assimetricamente para os extremos, pois nos dois limites finais do eixo x a curva do erro supera a da informação e, portanto, não traz dados significativos para proficiências que estejam posicionados nesses locais. Desse modo, a partir de uma rápida inspeção visual é possível determinar que a representação da direita apresenta maior quantidade de informação do que a da esquerda. Registra-se também como evidências expressivas nesse sentido a E09, E10, E11, E12 e E13 (escolhidas por alcançarem mais de duas unidades de referência no eixo y).

Embora ao analisar que, aparentemente, as CCI das evidências E01, E02, E03, E04, E06, E07, E14, E15, E16, E17, E18, E19, E20, E27, E28, E29 e E32 sejam mais tímidas, entende-se que todas apresentam algum nível de informação, especialmente por apresentarem parâmetro “a” superior a 0,7. Exceto a E19 que, como mencionado, não apresentou boa qualidade em se tratando de discriminação e, por esse motivo, foi descartada. No Apêndice C estão detalhadas as CCI e CII de todas as evidências para consulta mais aprofundada.

Prosseguindo, o somatório das informações pode ser demonstrado no Gráfico 6. A linha contínua e azul diagramada na ilustração representa o comportamento da CIT, enquanto a linha pontilhada e vermelha o erro padrão da medida. Atenta-se para o fato que nas extremidades de θ , o teste fornece mais erro do que informação legítima, pois o erro padrão sobrepuja a curva de informação. Portanto, pode-se concluir que existe uma boa definição da curva, alertando que o constructo é indicado para medir estudantes com habilidade de média para baixa.

Gráfico 6 - CIT da prova prática do Curso Técnico em Mecânica



Fonte: elaborado pelo autor

Tanto as análises ancoradas na TCT quanto TRI, bem como os produtos gráficos gerados, permitiram manter todas as 32 evidências na composição dos resultados, mesmo que algumas trouxessem pouca informação, porém sem prejuízos estatísticos para o processo. A calibração se mostrou estável e atendendo critérios de qualidade previamente pesquisados na literatura, assentindo o avanço para a próxima fase que é a estimação das proficiências de cada estudante, detalhada na sequência.

4.3 ESTIMAÇÃO DAS PROFICIÊNCIAS

No processo avaliativo empregado, um valor numérico de habilidade é atribuído a cada indivíduo, isto é, mensura-se quanto de proficiência do domínio psicomotor um estudante conseguiu demonstrar no teste. Pasquali e Primi (2003) argumentam que a métrica do escore padrão é ancorada na média (0) e que na prática a grande maioria destes resultados pertencem ao intervalo de -3 a +3, isto porque numa curva de

distribuição normal entre esses pontos situam-se 99,97% de todos os sujeitos da população.

No âmbito deste estudo, observou-se o comportamento isolado de cada um dos 1.498 respondentes com base na parametrização realizada na fase 2 do Bilog-MG, preferindo realizar as análises na escala original gerada pelo programa (0, 1). O aluno com maior proficiência atingiu θ de 1,58, enquanto o com o menor desempenho registrou apenas -3,35. Percebe-se aqui duas características importantes: a) a avaliação prática empregada não é ideal para aferir com precisão essa amostragem, pois grande parte do poder de discriminação está entre -2,0 a -0,5, enquanto o conjunto de dados apresenta amplitude de -3,35 e 1,58; e b) a amostra também extrapolou a faixa mencionada por Pasquali e Primi (2003), apresentando indivíduos com traço latente superior a três desvios padrões, embora não necessariamente indique um problema, mas sim um ponto de atenção.

Observando os dados destacados na Figura 13, além dos limites já descritos, utilizou-se de outras estatísticas para maior compreensão. Segmentando por quartis, 25% da amostra está abaixo da proficiência -0,45 e 75% dos estudantes possuem habilidade estimada de até 0,66. O valor que representa exatamente o meio dessa proporção, também denominado de mediana, é de 0,01, além de apenas 10% do conjunto possuir valores mais elevados de proficiência, entre 1,27 e 1,58. Ao comparar também as medidas de tendência central e variância do pressuposto (0, 1) com a realidade, verificou-se congruência, uma vez que a média encontrada corresponde a 0,012 e o desvio padrão a 0,927, estando dentro dos limites aceitáveis.

Figura 13 - Distribuição das proficiências dos estudantes avaliados

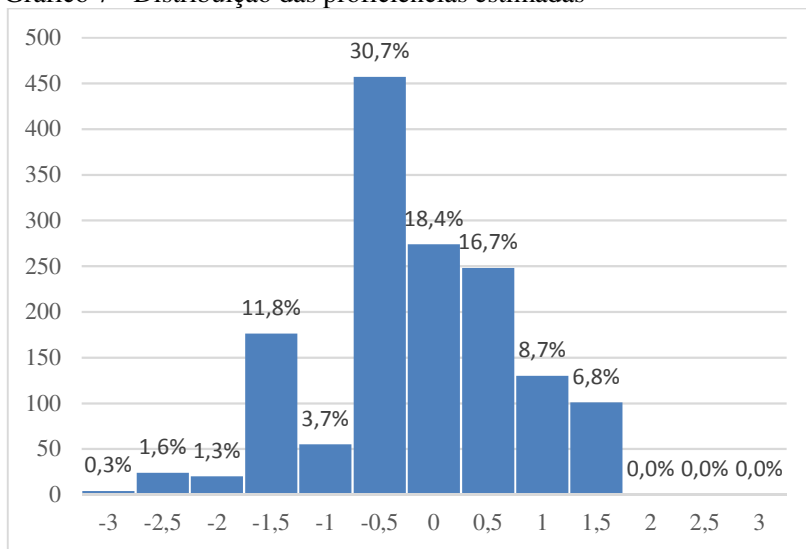


Fonte: elaborado pelo autor

Outro modo de visualizar a distribuição do traço latente está no Gráfico 7, onde demonstra-se o cruzamento da habilidade com a sua frequência. Chama-se atenção para a maior concentração no ponto “-0,5”, com 30,7% da amostra, seguido por “0” com 18,4% e “0,5” com 16,7%.

Somados, o intervalo (-0,5 a 0,5) agrupa 65,7% do total de estudantes avaliados. Além disso, a partir de -2 de proficiência não houve registros, enquanto na outra extremidade obteve-se valores próximos da unidade.

Gráfico 7 - Distribuição das proficiências estimadas



Fonte: elaborado pelo autor

Após os processos de calibração das evidências e estimação das proficiências, buscou-se realizar uma análise para validação do instrumento utilizado, conforme será explanado na sequência.

4.4 ANÁLISE DO INSTRUMENTO DE AVALIAÇÃO

Conforme apontado por Pasquali (1997) são duas as características essenciais para credibilidade de um teste: validade e fidedignidade. Sendo assim, debruçou-se em verificar esses dois pontos com vistas a meta-avaliação, ponderando a qualidade da prova prática aplicada aos estudantes do Curso Técnico em Mecânica.

Na dimensão “validade” o foco era verificar se o instrumento mensurava supostamente o nível de proficiência dos indivíduos no domínio psicomotor da área avaliada, considerando as estratégias de validação de conteúdo e de constructo. No primeiro caso, embora as evidências testadas tenham sido elaboradas por docentes com domínio técnico na área, verificou-se a discriminação de cada uma delas quanto à

coerência no conjunto, além de uma reinterpretação por grupo de especialistas a posteriori, sem contato com as etapas de preparação e aplicação do teste.

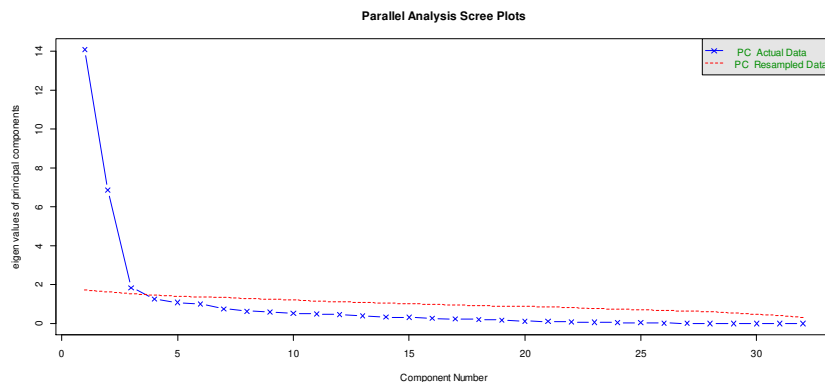
Observando os parâmetros “a”, percebeu-se uma amplitude que varia entre 0,580 e 5,830. Conforme comprovado na Tabela 4, apenas uma única evidência demonstrou comportamento insatisfatório (discriminação baixa), enquanto as demais (96,9%) foram plenamente satisfatórios. De todo modo, colocou-se a E19 (“a” = 0,580) para apreciação técnica dos especialistas sob o prisma da qualidade da informação trazida, questionando se agregava na interpretação dos resultados ou se seria considerada apenas para balizamento das demais. Na opinião dos participantes (análise de juízes), todas as evidências apresentam algum nível de informação e, portanto, também são relevantes, porém nesta em particular inferiu-se cenários onde por motivos de disponibilização da infraestrutura ou didática dos próprios docentes, o comportamento de resposta dos estudantes poderia ser inconsistente. Logo, além da informação de classificação estatística, a visão dos peritos confirmou a validade de conteúdo já que o conjunto daquilo que foi solicitado ao estudante é coerente com o domínio psicomotor demandado para a realização das atribuições profissionais do Técnico em Mecânica.

Para validação do constructo observou-se a curva de informação do teste e se havia ou não unidimensionalidade, ou seja, se o domínio psicomotor era o fator mais influente no teste e se conseguiria explicar mais de 50% da variância dos resultados. A CIT, já apresentada no Gráfico 6, demonstrou que este instrumento é adequado apenas para estudantes com proficiência entre -2,0 a -0,5, ou seja, aqueles que se encontrariam abaixo da média e, portanto, não estando adequado para investigação do nível de aptidão generalizado de qualquer estudante. Todavia, isso não desabona sua validação, apenas denota a carência de evidências mais complexas e que mobilizam maior proficiência dos discentes.

No que tange a verificação da dimensionalidade, realizou-se a análise fatorial por correlações tetracóricas. Conforme visualizado no Gráfico 8, há uma dimensão preponderante, porém existe a sinalização de presença de uma segunda dimensão. Embora não haja uma diretriz consolidada entre os psicometristas quanto ao percentual de explicação das soluções fatoriais, em geral os autores apontam para 50% do total. Contudo, O’Grady (1982) alerta sobre as especificidades de cada estudo e sobre os vieses relacionado ao tamanho da amostra e quantidade de itens do instrumento, onde estes valores podem gerar a sub ou superestimação

dos fatores. Nesse sentido, considerando que a proporção da primeira dimensão explica 48,1% e, portanto, muito próxima aos 50%, segue-se o ajuizamento de Abelson (1995) quanto a avaliação da medida ser ponderada analisando a concordância entre a magnitude do delineamento teórico e metodológico com as descobertas empíricas. Neste caso, considerando as características do constructo e do instrumento, assume-se que a unidimensionalidade foi suficientemente relevante para desdobramento das análises TCT e TRI.

Gráfico 8 - Análise da dimensionalidade do instrumento



Fonte: elaborado pelo autor

Na Tabela 5 também buscou-se comparar as cargas fatoriais considerando a unidimensionalidade e bidimensionalidade do modelo logístico de dois parâmetros, uma vez que este foi escolhido para estimação dos parâmetros das evidências e proficiência dos indivíduos desta pesquisa. Estes valores podem estar entre -1,00 e +1,00, onde o 0 indica a total ausência de covariância entre a variável e o fator, e resultados superiores a 0,32 qualificam-no como representante útil, já que corresponde a 10% da variância explicada (TABACHNICK e FIDELL, 1996).

Tabela 5 - Cargas fatorias comparando modelo unidimensional e bidimensional

Ev	Unid.	Bid.		Ev	Unid.	Bid.	
	F1	F1	F2		F1	F1	F2
1	0,692	0,084	0,734	17	0,428	0,428	0,185
02	0,581	0,045	0,630	18	0,394	0,427	0,165
03	0,708	0,002	0,779	19	0,311	0,363	0,074
04	0,658	0,070	0,758	20	0,627	0,819	0,130
05	0,823	0,028	0,833	21	0,649	0,949	0,218
06	0,601	0,115	0,656	22	0,731	0,786	0,326
07	0,531	0,144	0,514	23	0,702	0,776	0,292
08	0,942	0,006	0,958	24	0,642	0,878	0,042
09	0,963	0,003	0,972	25	0,681	0,857	0,106
10	0,974	0,067	0,994	26	0,707	0,995	0,184
11	0,963	0,021	0,978	27	0,598	0,864	0,109
12	0,893	0,063	0,942	28	0,669	0,956	0,183
13	0,944	0,026	0,967	29	0,633	0,767	0,206
14	0,530	0,432	0,343	30	0,643	0,759	0,178
15	0,450	0,445	0,229	31	0,689	0,753	0,314
16	0,601	0,629	0,175	32	0,590	0,842	0,046

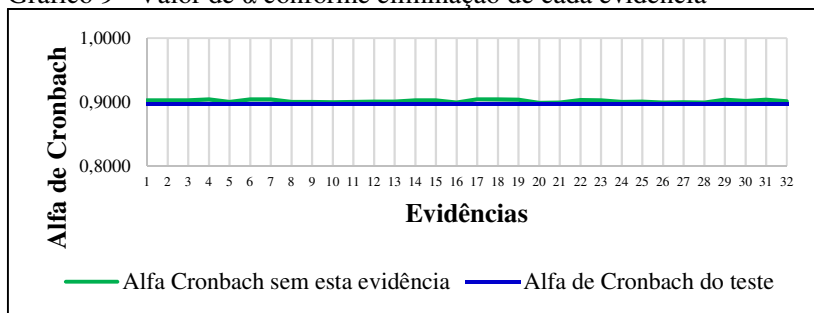
Fonte: elaborado pelo autor

Percebe-se que, quando aplicável uma única dimensão, praticamente todos os critérios atendem este postulado, com exceção da E19 (que posteriormente indicará outros aspectos para sua exclusão), justificando também a modelagem unidimensional escolhida como método. Todavia, o modelo bidimensional é um ponto de atenção para futuros trabalhos, pois até a E13 a F2 é mais dominante, enquanto as demais evidências apontam maior importância para F1. Uma possível inferência sobre esse comportamento é que as treze primeiras estão relacionadas a apenas uma tarefa das três atividades que os estudantes deveriam desenvolver “realizar um desenho técnico em meio eletrônico”, demonstrando que há um comportamento diferenciado em relação ao domínio psicomotor neste caso em relação aos demais parâmetros de avaliação. Outra possibilidade que se ventila é a indissociabilidade desta atividade entre os domínios cognitivo e psicomotor, destacando que ali não foi avaliado com exclusividade o domínio que pretendia-se. Embora do ponto de vista estatístico essa análise seja relevante, na prática os comportamentos estão relacionados com as demandas subsequentes, não

fazendo sentido numa apreciação dos especialistas destacar em duas dimensões distintas o constructo avaliado.

Já na dimensão “fidedignidade”, a principal ferramenta utilizada para avaliação da consistência interna do teste foi o Alfa de Cronbach, que calcula um indicador a partir da variância de cada uma das evidências individualmente e da variância total, permitindo identificar se há ou não confiabilidade. Com base nos dados parametrizados, verificou-se que o coeficiente da prova em particular é de 0,8974 e, apoiando-se na referência de Nunally (1978) e Martins e Theóphilo (2009), devido ao valor ser superior a 0,70 pode-se concluir sobre a credibilidade desta medida. Além disso, foi analisado também a exclusão de cada item e seu impacto no valor de α , conforme representado no Gráfico 9.

Gráfico 9 - Valor de α conforme eliminação de cada evidência



Fonte: elaborado pelo autor

Interpretando, a linha azul deixa pré-fixado o valor de 0,8974 como critério balizador para comparação, representando, portanto, o valor de α relativo ao teste no geral. Já a linha verde denota o cálculo aplicado desconsiderando a evidência caracterizada no eixo da abscissa, oscilando conforme a influência da sua importância para o conjunto. Percebe-se que na completude não há destaques que sinalizem problema em nenhum dos pontos do gráfico, pois os valores variaram entre 0,8985 e 0,9044, flutuação essa de +0,0012 a +0,0070 de incremento no coeficiente geral, considerado, neste âmbito, praticamente imperceptível. Mesmo a E19 que transpareceu dificuldades em relação à discriminação, quando suprimida gerou 0,9036 de resultado.

Além do coeficiente de precisão apresentado, outro modo de verificar a fidedignidade é utilizando-se a própria CIT, calculada por meio da FIT, e que considera a soma de todas as informações fornecidas por cada evidência. Em conformidade com o Gráfico 6, percebe-se que ao

longo dos diferentes níveis de θ há valores que ultrapassam 40, enquanto o critério de referência idealizado para garantir uma boa precisão da medida psicométrica é 10 (HAMBLETON, 2004). O intervalo que expressa essa condição em sua plenitude é delimitado, aproximadamente, entre -2,2 a -0,4.

Consubstanciado por essas análises de validade e fidedignidade, confirmou-se a viabilidade de avançar na construção de uma escala de proficiência que retratasse o comportamento dos estudantes frente a mobilização do domínio psicomotor, descrito na seção seguinte.

4.5 CONSTRUÇÃO DA ESCALA

Em posse dos resultados da estimação dos parâmetros das evidências e das proficiências dos estudantes calculadas na mesma métrica, é possível estabelecer uma escala padrão para mensuração do domínio psicomotor. Devido aos pressupostos metodológicos da TRI e da programação do software utilizado, inicialmente tem-se média 0 (zero) e desvio padrão 1 (um) para as proficiências (θ), porém como estes valores são de difícil compreensão pela comunidade escolar, optou-se por realizar uma transformação linear de representação numérica, respeitando sua proporcionalidade em relação a referência original.

Esta conversão foi praticada por meio das equações (3) e (4) (descritas na seção 3.9), reposicionando para a escala (100, 10), onde a média nesse novo formato é 100 e o desvio padrão 10 pontos. A motivação desta escolha teve caráter arbitrário, selecionando valores que facilitassem a sua visualização e permitissem a posterior associação com os conteúdos existentes em cada critério avaliativo, possibilitando assim a interpretação pedagógica nos diferentes indicadores de proficiência. O procedimento adotado que propiciou esse entendimento foi a identificação de níveis e itens âncoras fundamentadas por análises estatísticas e o posicionamento das evidências por especialistas com domínio técnico na área avaliada.

Precedentemente, estimou-se a probabilidade de atendimento da evidência (“sim”) em cada um dos níveis âncoras e na perspectiva desta pesquisa decidiu-se por apresentar essas importâncias a cada meio desvio padrão (ver Apêndice B). Contatou-se a existência de 7 níveis âncora (75, 80, 85, 90, 95, 100 e 105) e, segundo as condições adotadas, não houve evidências posicionadas em valores inferiores ou superiores aos limites desta faixa. Esse intervalo também indica que esta é a região da escala onde há o maior quantitativo de informação, oportunizando identificar para cada uma das etapas de progresso o que os estudantes realmente são

capazes de fazer. Cabe destacar que estes foram definidos a partir da lotação de cada evidência na graduação métrica estabelecida, metodologicamente operacionalizada em dois estágios distintos: a) posicionamento por critério estatístico e b) posicionamento por critério técnico.

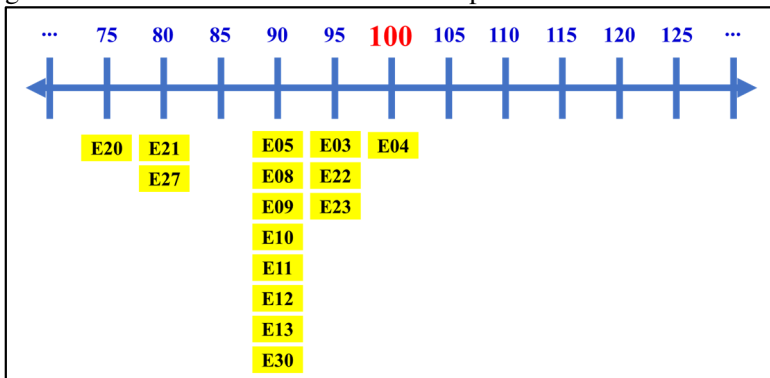
O critério estatístico foi uma reinterpretação dos três preceitos estabelecidos por Andrade, Tavares e Valle (2000) para definição de itens âncora, pois tais equações consideram também a proporção de acerto ao acaso (parâmetro “c”), não se aplicando em sua plenitude nessa pesquisa. Tendo em vista isso, refletiu-se sobre novas expressões que representassem melhor o ponto na escala de proficiência que de fato representasse o momento em que o traço latente fosse característico na amostra, sem que no nível imediatamente anterior houvesse uma probabilidade acentuada. Portanto, estabeleceu-se o seguinte regulamento para ancoragem nesse escopo:

$$P(U = 1 | \theta = Z) \geq 0,60 \quad (6)$$

$$P(U = 1 | \theta = Y) < 0,50 \quad (7)$$

Isso significa dizer que o local em que as evidências estão lotadas na escala, nessa perspectiva, é o nível em que pelo menos 60% dos estudantes atendam o estabelecido para atribuir marcação “sim” e menos da metade (50%) tem este mesmo desempenho no nível imediatamente anterior (ver destaques em amarelo no Apêndice D). Apura-se que das 32 evidências apenas 15 respeitaram esse normativo, conforme diagramação da Figura 14.

Figura 14 - Posicionamento das evidências pelo critério estatístico



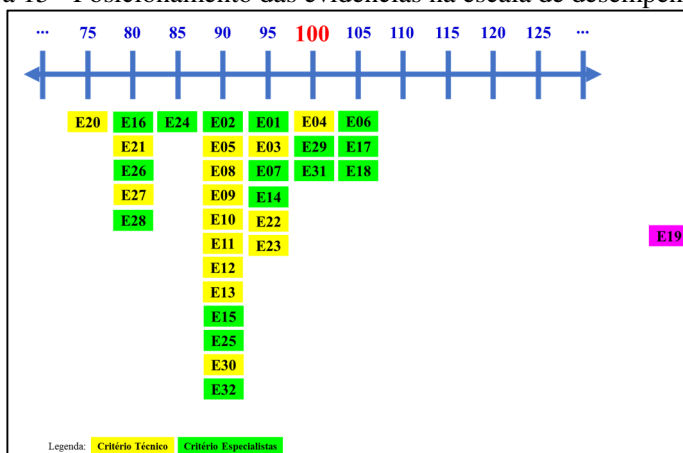
Fonte: elaborado pelo autor

Conforme destaque da imagem, percebe-se que todas ficaram situadas em valores iguais ou inferiores à média (100), indicando a ausência de elementos investigativos para estudantes com proficiências mais elevadas. Além disso, cada nível está desdobrado pelos respectivos códigos de referência, sendo o 90 com a maior quantidade de evidências (8 no total), seguido pelo 95 (3) e 80 (2). Os pontos 75 e 100 tiveram apenas uma, enquanto não houve elementos para o 85 e 105 nesta etapa.

As 17 evidências que não foi possível de ser ancoradas nessa abordagem tiveram encaminhamento para a segunda fase de posicionamento por critério técnico. Este processo se caracterizou pela apreciação de especialistas (na área avaliada), sobre o melhor ponto de lotação conforme o conjunto já pré-fixado na etapa anterior. Inicialmente houve uma análise prévia das evidências já fixadas e os peritos tiveram boa receptividade, visto que a disposição ao longo da régua de pontuação era gradativa e fazia sentido do ponto de vista pedagógico de evolução de complexidade.

Além disso, apresentava-se sempre duas opções para tomada de decisão, escolhas estas realizadas pelo nível em que os estudantes atingiam pelo menos 60% de probabilidade de acerto e o nível antecessor, independentemente do valor. Por exemplo, para a E01 foi disponibilizada as alternativas 95 (0,71) ou 90 (0,52) e por definição dos peritos sua acomodação ficou no nível 95, conforme harmonização de sentido com as demais já estipuladas. A Figura 15 consolida o resultado cumulativo das duas etapas.

Figura 15 - Posicionamento das evidências na escala de desempenho



Fonte: elaborado pelo autor

Alguns pontos interessantes a serem observados: em amarelo estão representadas as evidências já trazidas da etapa anterior; em verde as escolhas realizadas pelos especialistas para ancoragem no critério técnico; e em roxo o caso desconsiderado para fins de posicionamento. Particularmente, sobre esta última circunstância, a decisão foi determinada pelo parâmetro de discriminação ser baixo e pelas constatações apuradas na opinião dos peritos, uma vez que a infraestrutura necessária para sua concretização é de comportamento “aleatório” pelas escolas, não sendo um padrão dos laboratórios de usinagem a utilização de rugosímetros.

Percebe-se também que após esse processo, o nível 85 teve ancoragem, assim como o 105. O ponto 90 continuou sendo o que agrupou maior número de evidências, chegando a representar 37,5% do total (12 das 32). Outra análise possível é a correspondência com a CIT, pois apenas houve três posicionamentos acima da média, ratificando que o instrumento não apresenta muitos critérios avaliativos para mensurar estudantes com proficiências mais elevadas do domínio psicomotor.

Distribuídas as evidências conforme alocação em seus respectivos níveis, podemos passar a buscar possíveis interpretações pedagógicas que auxiliem a entender o comportamento dos estudantes nos diferentes pontos da escala, bem como estabelecer uma classificação pedagógica que faça sentido para estabelecer análises gerenciais, passo esse que se desvela a seguir.

4.6 INTERPRETAÇÃO PEDAGÓGICA

A penúltima etapa da análise dos resultados culminou com a interpretação da escala criada à luz dos preceitos pedagógicos, para que docentes e administradores escolares pudessem utilizá-la de modo a oportunizar melhorias nos processos de ensino-aprendizagem. Nesse sentido, o crivo técnico estabelecido por especialistas da área é fundamental, pois cada área tecnológica da formação profissional possui seu universo e linguagem própria. A preocupação aqui é traduzir os resultados estatísticos em informações relevantes e que possam retratar o comportamento dos diferentes níveis de desempenho, explicando como se manifesta a evolução da proficiência do domínio psicomotor para os discentes do curso técnico em mecânica.

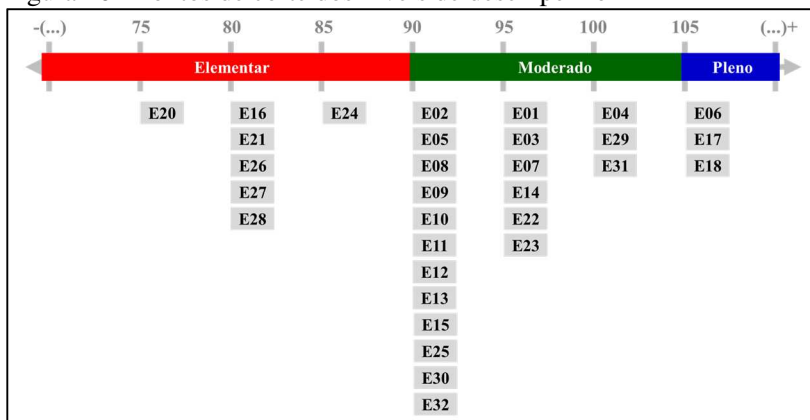
Para iniciar essa tarefa, os peritos foram convidados a analisar o resultado do posicionamento das evidências (demonstrado na Figura 15). Em cada nível da escala há um agrupamento de itens, seja este por critério estatístico ou técnico, e que no conjunto delineiam afinidades para

composição do mesmo patamar. Na visão dos convidados, fez sentido a proximidade entre elas e a sua hierarquia quantitativa, permitindo indicar desde os critérios mais fáceis até as mais difíceis, naturalmente dentro da amplitude dos parâmetros já conhecidos.

Após essa etapa de apreciação, foi solicitado para que avaliassem as faixas de desempenho, isso é, a definição de cortes em determinados pontos da escala que retratassem comportamentos de proficiência por contiguidade. Num primeiro momento deixou-se aberto para livre associação e posteriormente necessitou-se direcionar para que fossem estabelecidos de três a cinco segmentações, visando a utilização gerencial das informações. A opção foi pela utilização de três níveis de proficiência, visto também não haver uma grande variabilidade de ancoragem.

Para cada um deles foi discutida uma nomenclatura que expressasse não somente a progressão da proficiência, mas também sua clareza, assertividade e discriminação. Outro ponto adicional nessa discussão estava relacionado ao cuidado para não se confundir com as terminologias já utilizadas na avaliação objetiva (“abaixo do básico”, “básico”, “adequado” ou “avançado”), para não unificar classificações de traços latentes distintos. Desse modo, elencou-se o sequenciamento: “elementar”, “moderado” e “pleno”, ordenados de modo cumulativo e progressivo, para descrever as tipificações admitidas nesse contexto. O resultado é revelado na Figura 16.

Figura 16 - Pontos de corte dos níveis de desempenho



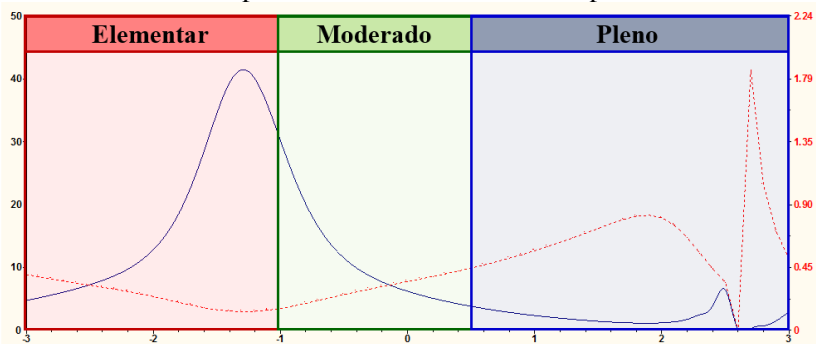
Fonte: elaborado pelo autor

Observa-se que na camada “elementar”, os estudantes provavelmente conseguem responder de modo satisfatório as evidências:

E16, E20, E21, E24, E26, E27 e E28 caracterizando 22,6% do conjunto total. O maior agrupamento deu-se no campo “moderado”, composto pelas E01, E02, E03, E04, E05, E07, E08, E09, E10, E11, E12, E13, E14, E15, E22, E23, E25, E29, E30, E31 e E32, compondo 67,7% daquilo que foi avaliado. No nível “pleno”, somente houve três ancoragens: E06, E17 e E18, representando 9,7%. Reforça-se que a E19 não foi contemplada no estudo interpretativo pelo seu baixo poder de discriminação.

Além disso, ao analisar os níveis de desempenho na CIT, percebeu-se que a maior quantidade de informação está na faixa “elementar”. O intervalo “moderado” também é relevante, embora não tão expressivo quanto o anterior. Entretanto, o nível “pleno” tem pouco nível de informação permanecendo abaixo da linha tracejada do erro, indicando fragilidade do instrumento para avaliação dos estudantes nesse patamar, conforme é possível ser observado no gráfico a seguir.

Gráfico 10 - CIT comparada com os níveis de desempenho da escala



Fonte: elaborado pelo autor

Vale destacar também que, embora a CIT represente um teste de baixa a média complexidade, os especialistas técnicos optaram por definir um nível que destaque os alunos que alcancem as E06, E17 e E18, denominando de “pleno”. Na visão deles, os estudantes que conseguem desempenhar estes critérios já apresentam comportamento diferenciado dos demais. Nada impede também que, no futuro, com a inserção de evidências mais elevadas como, por exemplo, “usinou o comprimento total da peça conforme tolerância de $\pm 0,05$ ”, estes cortes sejam redimensionados e ampliados para mais uma categoria, a qual sugere-se “elevado”.

Desse modo, visando interpretar melhor o que cada um desses níveis representa, pensou-se numa descrição sumária relativa ao que os

discentes naquele estágio eram provavelmente capazes de fazer. Esta síntese somente foi possível porque os especialistas levaram em consideração o próprio conteúdo de cada evidência e seus agrupamentos por faixa. O Quadro 6 consolida esse entendimento.

Quadro 6 - Descrição pedagógica sumária de cada nível de desempenho

Nível (recorte da proficiência)	Descrição
Elementar (menor de 90)	O estudante neste nível provavelmente é capaz de aplicar procedimentos de saúde e segurança nas atividades profissionais sob sua responsabilidade, utilizar instrumentos de medição/calibração associados à área mecânica e manipular dispositivos para montagem e usinagem mecânica.
Moderado (maior ou igual 90 e menor que 105)	O estudante neste nível, além do descritivo anterior, provavelmente é capaz de aplicar representações gráficas de peças mecânicas em software de desenho assistido por computador, realizar cálculos de parâmetros de cortes para peças e executar procedimentos de montagem e usinagem mecânica.
Pleno (maior ou igual de 105)	O estudante neste nível, além dos descritivos anteriores, provavelmente é capaz de executar tarefas de usinagem com maior precisão, observando parâmetros de corte e tolerâncias específicas.

Fonte: elaborado pelo autor

Avançando nas discussões, buscou-se um segundo nível de interpretação, agora não mais sintético, e sim com maior aprofundamento e comparação com as capacidades previstas na matriz de referência. Contudo, percebeu-se que o método utilizado para descrição dessas habilidades estava ancorado na Taxonomia de Bloom Revisada e, portanto, fazia menção ao domínio cognitivo exclusivamente. Foi proposto aos especialistas técnicos que revisitassem cada uma das evidências e realizassem sua compreensão pelo domínio psicomotor, isto é, pensando na dimensão do “saber fazer”. Desse modo, procedeu-se com novos agrupamentos que retratassem melhor o traço latente aqui avaliado na perspectiva dos estudantes do curso técnico em mecânica, conforme desdobrado no Quadro 7.

Quadro 7 - Correlação das capacidades do domínio cognitivo (matriz de referência) x domínio psicomotor (nova proposta)

Ev.	Capacidades do domínio afetivo	Capacidades do domínio psicomotor					
		A	B	C	D	E	F
E1	C1	x					
E2	C8	x					
E3	C8	x					
E4	C4	x					
E5	C1	x					
E6	C4	x					
E7	C1	x					
E8	C8	x					
E9	C8	x					
E10	C8	x					
E11	C8	x					
E12	C8	x					
E13	C8	x					
E14	C1		x				
E15	C2				x		
E16	C3		x				
E17	C5				x		
E18	C5				x		
E19	C8				x		
E20	C8						x
E21	C8						x
E22	C1			x			
E23	C1			x			
E24	C3					x	
E25	C3		x				
E26	C8						x
E27	C8						x
E28	C8						x
E29	C5		x				
E30	C5		x				
E31	C1		x				
E32	C4					x	
Total		13	6	2	4	2	5

Fonte: elaborado pelo autor

Utilizou-se letras como nomenclatura com o objetivo de não comparabilidade linear com as capacidades do domínio cognitivo, onde:

- A. Manipular ferramentas de desenho assistido por computador para representação de peças mecânicas
- B. Utilizar instrumentos de metrologia para medições utilizadas no processo de usinagem
- C. Executar comandos manuais para configuração de máquinas mecânicas
- D. Usinar peças mecânicas utilizando ferramentas específicas
- E. Montar peças mecânicas por meio de fixação
- F. Demonstrar procedimentos de segurança do trabalho na realização das atividades

Com base nesse novo arranjo comparou-se a proposta com as categorias da dimensão da taxonomia proposta por Dave (1970) e Simpson (1972) e a Pirâmide de Miller (1990). O resultado foi fruto de discussão com os peritos e pode ser demonstrado no Quadro 8.

Quadro 8 - Classificação das capacidades psicomotoras

Capacidade	Classificação Taxonomia	Pirâmide de Miller
A	Nível 4	4. Fazer
B	Nível 3	3. Mostrar como faz
C	Nível 4	4. Fazer
D	Nível 5	4. Fazer
E	Nível 4	4. Fazer
F	Nível 3	3. Mostrar como faz

Fonte: elaborado pelo autor

As capacidades “B” e “F” podem ser consideradas como pertencentes ao terceiro estágio da Pirâmide de Miller, isto é, estão associadas a um processo de demonstração de como fazer as atividades. Por sua vez, o nível esperado de comportamento é o “automatismo” (nível 3) conforme a taxonomia do domínio psicomotor, pois entende-se que sua mobilização seja de modo intuitivo e sem muito esforço motor. Já “A”, “C” e “E” pertencem ao quarto extrato da pirâmide, assumindo, portanto, os níveis mais elevados e complexos de avaliação segundo a teoria de Miller, pois focam no “fazer”. Observando sua classificação, diz respeito

às “respostas complexas” (nível 4), já que demandam comportamento voluntário e reflexos, apresentando respostas a estímulos motores. Por fim, embora “D” esteja na mesma condição das três anteriores, a conduta já é de nível 5 “adaptação”, devido a necessidade de adequação em situações novas e adversas conforme a atividade está sendo realizada.

Dessa forma, considerando o novo formato de descrição das capacidades psicomotoras e a metodologia apresentada por Mager (1979) para a formulação de um objetivo educacional específico, foi possível ampliar a interpretação da escala de desempenho, conforme diagramado no Quadro 9, trazendo mais informação pedagógica para permitir ações e intervenções na progressão da proficiência dos estudantes.

Quadro 9 - Descrição pedagógica detalhada dos níveis de desempenho

Nível de desempenho: Elementar
<p>Os estudantes neste nível provavelmente são capazes de:</p> <p>B - Utilizar instrumentos de metrologia para medições utilizadas no processo de usinagem conforme especificações do projeto, considerando orientações do fabricante.</p> <p>E - Montar peças mecânicas por meio de fixação conforme seleção de dispositivo para montagem, considerando materiais e ferramentas para a realização da manutenção.</p> <p>F - Demonstrar procedimentos de segurança do trabalho na realização das atividades conforme</p> <ul style="list-style-type: none">• utilização dos EPI, considerando a operação e manutenção de máquinas e equipamentos;• organização do ambiente de trabalho, considerando as atividades de manutenção mecânica de máquinas e equipamentos;• preservação da segurança e integridade pessoal, considerando os processos de substituição de peças e componentes mecânicos.

Quadro 9 - Descrição pedagógica detalhada dos níveis de desempenho (continuação)

Nível de desempenho: Moderado
<p>Os estudantes neste nível, além do descrito anteriormente, provavelmente são capazes de:</p> <p>A - Manipular ferramentas de desenho assistido por computador para representação de peças mecânicas conforme</p> <ul style="list-style-type: none">• utilização dos tipos de linhas, considerando normas técnicas de desenho; <p>corpo central da peça, considerando o desenho do croqui;</p> <ul style="list-style-type: none">• preenchimento da legenda, considerando as informações de autoria, escala e detalhamento dos materiais;• vistas ortográficas da peça mecânica, considerando o desenho do croqui;• cotas da peça mecânica, considerando o desenho do croqui;• furo roscado, considerando o desenho do croqui;• canal, considerando detalhes do projeto mecânico;• diferentes tipos de rebaixo, considerando detalhes do projeto mecânico. <p>B - Utilizar instrumentos de metrologia para medições utilizadas no processo de usinagem conforme</p> <ul style="list-style-type: none">• registro dos parâmetros de corte, considerando o material e ferramenta especificados em ordem de serviço;• registro de dimensões limites do eixo, considerando tabela do sistema de tolerância e ajustes especificados em ordem de serviço;• dimensão do alojamento, considerando a realização da manutenção;• registro das dimensões real do eixo e do rolamento, considerando especificações em ordem de serviço. <p>C - Executar comandos manuais para configuração de máquinas mecânicas conforme</p> <ul style="list-style-type: none">• desvios permissíveis, considerando tabela de sistemas de tolerância e ajustes;• classe de tolerância, considerando tabela técnica de montagem de rolamentos. <p>D - Usinar peças mecânicas utilizando ferramentas específicas conforme sequência das etapas estabelecidas, considerando especificações técnicas em ordem de serviço.</p> <p>E - Montar peças mecânicas por meio de fixação conforme encaixe de anel interno do rolamento em face de encosto do eixo, considerando as especificações técnicas estabelecidas na ordem de serviço.</p>

Quadro 9 - Descrição pedagógica detalhada dos níveis de desempenho (continuação)

Nível de desempenho: Pleno
Os estudantes neste nível, além do descrito anteriormente, provavelmente são capazes de: A - Manipular ferramentas de desenho assistido por computador para representação de peças mecânicas conforme furos de centro, considerando detalhes do projeto mecânico. D - Usinar peças mecânicas utilizando ferramentas específicas conforme comprimento e diâmetro do rebaixo, considerando especificações técnicas do projeto.

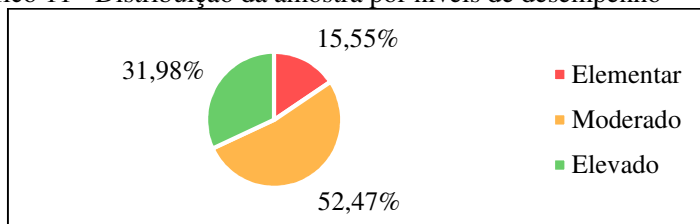
Fonte: elaborado pelo autor

A representação acima auxilia os docentes na compreensão de quais capacidades estão envolvidas em cada nível de desempenho, além da evolução da sua complexidade. Percebe-se que dos estágios “elementar” para “moderado” as habilidades “B” e “E” se repetem, porém com critérios de avaliação e condições diferentes. O mesmo acontece do “moderado” para o “pleno”, só que agora com “A” e “D”, permitindo observar os graus mais avançados possíveis de serem medidos nesse instrumento. Outro ponto a ser destacado é que a capacidade “F” apenas aparece no nível “elementar”, não havendo evidências de maior complexidade nesse âmbito.

4.7 APLICAÇÃO DA ESCALA DE DESEMPENHO

Considerando os níveis de desempenho propostos pelos especialistas, pôde-se classificar a amostra avaliada, conforme representação do Gráfico 11. A maioria dos estudantes pertence ao estágio “moderado”, com 52,47% do total, seguido por 31,98% dos alunos no grupo “pleno” e somente 15,55% na categoria “elementar”.

Gráfico 11 - Distribuição da amostra por níveis de desempenho



Fonte: elaborado pelo autor

Cruzou-se os dados categorizados com as variáveis descritas na análise estatística descrita, para verificar o comportamento da amostra de acordo com a UF de origem (Tabela 6), sexo (Tabela 7) e idade (Tabela 8). De modo geral, as tabelas trazem a distribuição absoluta e relativa, tal como o somatório total e o percentual de estudantes considerados “proficiente” (% prof.), ou seja, o montante agrupado dos níveis “moderado” e “pleno”, pois na visão dos especialistas técnicos os estudantes nesse intervalo já apresentam nível adequado do domínio psicomotor para “fazer” as atividades profissionais no mercado de trabalho.

Tabela 6 - Distribuição dos níveis de desempenho da escala por UF

UF	Elementar		Moderado		Pleno		Total	% Prof.
AM	0	0,00%	10	40,00%	15	60,00%	25	100,00%
BA	31	75,61%	9	21,95%	1	2,44%	41	24,39%
CE	8	47,06%	9	52,94%	0	0,00%	17	52,94%
ES	22	13,58%	78	48,15%	62	38,27%	162	86,42%
GO	12	15,58%	58	75,32%	7	9,09%	77	84,42%
MG	25	9,77%	146	57,03%	85	33,20%	256	90,23%
MS	0	0,00%	14	51,85%	13	48,15%	27	100,00%
MT	4	22,22%	8	44,44%	6	33,33%	18	77,78%
PE	1	2,94%	22	64,71%	11	32,35%	34	97,06%
PR	33	22,92%	76	52,78%	35	24,31%	144	77,08%
RJ	22	9,73%	132	58,41%	72	31,86%	226	90,27%
RS	15	21,43%	33	47,14%	22	31,43%	70	78,57%
SC	47	25,41%	77	41,62%	61	32,97%	185	74,59%
SE	0	0,00%	13	76,47%	4	23,53%	17	100,00%
SP	12	6,22%	96	49,74%	85	44,04%	193	93,78%
TO	1	16,67%	5	83,33%	0	0,00%	6	83,33%
Total	233	15,55%	786	52,47%	479	31,98%	1498	84,45%

Fonte: elaborado pelo autor

Analisando os dados, percebe-se que três estados (AM, MS e SE) apresentaram 100% dos estudantes como proficiente, deste destaque extra para Amazonas que obteve o maior percentual relativo no nível “pleno”. Os menores resultados ficaram, respectivamente, para BA (24,39% proficientes) e CE (52,94%) representando os valores mais destoantes do restante das UF. Ao verificar por faixa, BA registrou o maior percentual em “elementar” (75,61%), de mesmo modo TO em “moderado” (83,33%) e AM no último nível conforme já descrito. É claro também que a

apreciação relativa aqui é um assunto sensível, uma vez que a distribuição nos estados varia de 6 a 256 estudantes.

Tabela 7 - Distribuição dos níveis de desempenho da escala por sexo

	Feminino		Masculino		Total	
Elementar	15	11,45%	218	15,95%	233	15,55%
Moderado	75	57,25%	711	52,01%	786	52,47%
Pleno	41	31,30%	438	32,04%	479	31,98%
Total	131	100,00%	1367	100,00%	1498	100,00%
% Prof.	116	88,55%	1149	84,05%	1265	84,45%

Fonte: elaborado pelo autor

Quando o assunto é comparação por sexo, constatou-se pouca variação no estágio “pleno” chegando apenas a 0,74% de diferença, porém nos demais níveis já há certo grau de disparidade. Enquanto 15,95% dos homens estão com proficiência “elementar”, somente 11,45% das mulheres se enquadram da mesma forma (-4,5%). O inverso ocorre na conjuntura subsequente (“moderado”), com 52,01% dos representantes do sexo masculino contra 57,25% do outro público (+5,24). Dessa forma, esse conjunto de dados revelou tendência de o grupo feminino ter desempenho ligeiramente superior ao do outro sexo.

Tabela 8 - Distribuição dos níveis de desempenho da escala por classes de idade

Classes	Elementar		Moderado		Pleno		Total	% Prof.
18 -- 26	45	13,65%	570	53,67%	347	32,67%	1062	86,35%
26 -- 34	60	21,05%	130	45,61%	95	33,33%	285	78,95%
34 -- 42	22	18,64%	65	55,08%	31	26,27%	118	81,36%
42 -- 50	5	19,23%	16	61,54%	5	19,23%	26	80,77%
50 -- 58	0	0,00%	4	80,00%	1	20,00%	5	100,00%
58 -- 63	1	50,00%	1	50,00%	0	0,00%	2	50,00%
Total	33	15,55%	786	52,47%	479	31,98%	1498	84,45%

Fonte: elaborado pelo autor

A última estratificação fica por conta das idades, revelando algumas estatísticas interessantes. Embora as faixas etárias da quarta classe em diante possuam poucos representantes, a categoria “50 |--58” contou com 100% de estudantes proficientes, sendo que entre “58 |--63” essa proporção é apenas a metade (apenas dois indivíduos). O primeiro recorte é que conta com maior participação (70,9% da amostra) e por esse

motivo revela comportamento muito semelhante a distribuição total. A categoria “26 |- 34” é o intervalo que acolhe o maior valor relativo no nível “pleno” (33,33%), embora seja o segundo maior percentual no nível “elementar” (21,05%).

Os extratos aqui descritos refletem a realidade global de todos os estudantes que compuseram a pesquisa. Contudo, é possível realizar esse mesmo detalhamento por unidade escolar, ou seja, cada uma delas pode analisar seus resultados, verificar o desenvolvimento das capacidades em seus estudantes e definir um plano de ação para melhoria do seu processo educacional. Essa é a principal finalidade de um método de interpretação de resultados, visto que ele só será útil quando compreendido pela comunidade escolar (coordenadores pedagógicos, docentes, supervisores, etc.) e quando aplicado para redirecionamento das práticas pedagógicas ou fortalecimento de ações exitosas.

5 CONSIDERAÇÕES FINAIS

O histórico estigmatizado da educação profissional fez com que até o presente momento poucas políticas públicas tenham sido desenvolvidas no Brasil em prol da sua promoção, em especial aquelas relacionadas a avaliação da qualidade e controle da oferta pelas instituições de ensino. Embora presente no PNE e com metas estabelecidas, há muito a se avançar para triplicar o volume de matrículas nessa modalidade, denotando um dos desafios mais emergentes da educação do nosso país. Imbuído dessa atmosfera, esse trabalho promove o debate para trazer luz à suas particularidades e acrescentar contribuições no que tange à avaliação da aprendizagem dos estudantes.

Esta constatação favoreceu a formulação da questão de pesquisa, de como elaborar uma escala que permitisse avaliar na prática o domínio psicomotor da aprendizagem, evidentemente associado ao contexto da educação profissional. Para isso, estabeleceu-se um objetivo geral que era a proposição da própria escala, utilizando-se como metodologia principal de análise dos resultados a TRI. Cinco foram os objetivos específicos elencados para desdobramento do percurso que seria trilhado para mapeamento de sete principais contribuições deste estudo.

Uma primeira contribuição foi o realce do domínio psicomotor como elemento fundamental para avaliação de competências profissionais. A perspectiva de Perrenoud (1999) demonstra que a mobilização de um conjunto de recursos é o que possibilita afirmar quando uma pessoa é competente ou não, somando-se ao olhar de Bloom (1983) sobre os três domínios – cognitivo, psicomotor e afetivo – como elementos para alcance disso. Embora os três sejam importantes, a educação propedêutica e todas as suas avaliações em larga escala, seja por conveniência ou limitação de recursos, vem abordado apenas a primeira dessas dimensões. Contudo, para o mundo do trabalho, mais do que “saber”, o profissional deverá “fazer”, isto é, colocar “a mão na massa” e conseguir resolver situações com diferentes graus de complexidade no âmbito de suas atribuições.

São inúmeros os trabalhos que se propõem a detalhar como a cognição se manifesta, mas poucos estudos que aprofundam a questão psicomotora. Nessa vertente, não foi possível localizar com precisão algum estudo que fizesse uso de taxonomia de objetivos educacionais aplicada na educação profissional. Ao realizar o levantamento do estado da arte entendeu-se como mais adequada a classificação proposta por Dave (1970) e Simpson (1972), bem como a Pirâmide de Miller (1990) como estratégias para guiar a definição das capacidades que seriam

avaliadas. Esta, portanto, caracteriza-se como a principal contribuição desta pesquisa.

Pela pesquisa bibliográfica realizada (RSL), identificou-se o levantamento de experiências prévias, configurando-se como a segunda contribuição. Somente 0,1% das publicações na área (5 entre 4.994), foram selecionadas como aderentes à delimitação aqui estabelecida. Percebeu-se também que todas são estrangeiras, sem destaque para produção científica nacional, o que também indica a iminência do assunto. As aplicações encontradas são predominantemente da dimensão cognitiva dos estudantes ou professores da educação profissional e pouco se nota a dimensão psicomotora, que poderia ser observada por testes práticos (apenas citados em um dos casos). Em alguns artigos até ficaram evidentes a realização das provas por meio de situações-problema, embora ainda não houvesse aplicação prática, buscavam simular no papel o modo como os estudantes deveriam se comportar em uma determinada circunstância. Por esse motivo, constatou-se a deficiência em trabalhar com situações hipotéticas e não reais, reforçando a necessidade de avaliar na prática se o estudante consegue desenvolver uma determinada competência.

A RSL também trouxe algumas inspirações para a definição do método proposto nessa pesquisa, destacando: a necessidade de submeter o estudante em situações mais próximas da realidade possível para investigar o domínio psicomotor, pois os testes de múltiplas escolhas não conseguem aferi-lo com precisão; validade de utilização da TRI como abordagem metodológica para interpretação dos resultados; a seleção do modelo logístico de dois parâmetros, uma vez que o acerto ao acaso não se faz muito presente em testes práticos; emprego de rubricas, graduais, para descrição dos níveis de desempenho que, embora não aplicado, trouxe um apontamento de estudo futuro. Dessa forma, conseguiu-se alcançar o primeiro objetivo específico.

A terceira contribuição foi caracterizar os diferentes instrumentos de avaliação utilizados pelo SENAI como uma tentativa de medir o domínio psicomotor. O teste baseava-se na metodologia de situações-problema, com contexto relacionado à área avaliada e execução em laboratórios didáticos, ambiente simulado ao da prática profissional. Além de ter procedimentos claros e definidos desde o planejamento até a aplicação, destaca-se a utilização de matriz de referência como ferramenta para trazer clareza aos objetivos educacionais que deveriam ser verificados. Entretanto, sua construção era para o teste cognitivo e não auxiliou muito na análise dos resultados, demandando novos arranjos e transcrição de capacidades para uma nova redação. Aqui tem-se uma lição

aprendida quanto a construção do constructo já na metodologia exposta anteriormente, em especial ao uso da teoria detalhada por Dave (1970) e Simpson (1972). Nessa etapa da pesquisa, portanto, foi atendido o seu segundo objetivo específico.

A quarta contribuição está relacionada ao êxito de aplicação do ML2P no conjunto de dados, comprovando que é possível utilizar a TRI como metodologia para calibração dos itens e estimação das proficiências dos estudantes. Reforça-se que, nesse âmbito, não se entende como relevante o parâmetro “c” – acerto ao acaso –, pois na resolução das atividades ou conseguia-se demonstrar o que estava sendo solicitado ou não, sem espaços para que o estudante tenha contato com outras fontes e tente alternativas suplementares para alcance do mesmo resultado. Importante esclarecer que a própria prova foi construída nessa perspectiva. Desse modo, a análise dos dados caracterizou-se como descritiva, diagnóstica e propositiva.

Outro ponto de atenção identificado nessa pesquisa foi a análise do instrumento quanto à validade e fidedignidade. Embora todos os coeficientes utilizados e analisados com esse propósito tenham se mostrado satisfatório, ressalta-se o estudo da dimensionalidade do teste. Ainda que os resultados tenham sido apresentados como unidimensionais, pela sua definição hipotética e mantida pelos critérios estatísticos empregados, há de se refletir com maior afinco sobre a viabilidade de aplicação de uma prova que pudesse mensurar as dimensões cognitivas e psicomotoras concomitantemente, mais aderente com a proposta de Miller. Essa questão se configura como a quinta contribuição e, conjuntamente com a anterior, permitem responder ao terceiro objetivo específico.

Uma sexta contribuição está no método utilizado para construção da escala e sua interpretação pedagógica. Isso porque houve dois momentos para maior acurácia do posicionamento real das evidências na régua graduada. Um puramente por meio de cálculos e pressupostos bem delimitados pela literatura (critério estatístico) e outro pela participação ativa de especialistas da área (critério técnico). Esse último permitiu que as decisões tomadas ao longo da construção fossem válidas e trouxessem uma carga de significado na visão pedagógica. Nesse interim, foram construídas duas descrições que permitiam a análise da progressão das proficiências dos estudantes, uma mais sumária e com finalidades gerenciais, e outra mais detalhada, permitindo o replanejamento das atividades utilizadas em sala de aula para desenvolvimento das capacidades mais sensíveis em cada nível de desempenho. Com isso, conseguiu-se atender o quarto objetivo específico.

A sétima e última contribuição destacada aqui é o diagnóstico dos estudantes da educação profissional por meio da aplicação da escala proposta nesse estudo. A maioria absoluta, 84,45% da amostra, apresenta-se proficiente e pronta para iniciar suas atividades no mercado de trabalho, enquanto a camada de 15,55% ainda necessita mais capacitação para desenvolver suas atividades plenamente. Num processo dessa envergadura, é possível denotar que os valores são relativamente altos quando comparados com a distribuição normal, pois sempre haverá um estrato em cada extremidade da escala, tanto para os níveis “elementar”, quanto para o “pleno”. Também há diferenças significativas quando observado o comportamento do grupo por UF, sexo e idade. Com estas constatações chega-se ao alcance do quinto objetivo específico.

É válido destacar ainda que o modelo construído de interpretação dos resultados poderá ser incorporado na avaliação institucional já realizada pela instituição de ensino objeto de estudo. A proposição, deste método, é um artefato para que o SENAI possa refletir sobre seu processo avaliativo e incorporar novas práticas respaldadas em subsídios científicos, beneficiando estudantes e docentes na melhoria contínua do “ensinar” e do “aprender”.

Uma outra recomendação é maximização dos resultados e ampliação da sua utilização no contexto escolar, abrangendo inclusive outras fontes de inspiração, como a utilização para identificação de talentos (alta proficiência) para participação em competições nacionais e internacionais, a possibilidade de testar profissionais já em atuação no mercado de trabalho para certificação de competências (certificação ocupacional), bem como a identificação de lacunas de atualização tecnológicas ou competências específicas em trabalhadores já habilitados tecnicamente (diplomados).

Somado às contribuições já mencionadas, este estudo promove também o avanço da ciência testando teorias educacionais já estabelecidas e aplicando-as em outros campos, observando seus impactos e resultados. Resume-se, então, que esta dissertação cumpriu com os seus cinco objetivos específicos e o seu objetivo geral, além de subsidiar discussões para futuras pesquisas como:

- Construção de uma matriz de referência para a educação profissional, com linguagem mais apropriada ao contexto dos docentes e ancorada na proposta de educação por competência.
- Aplicação de outras metodologias de validação do instrumento, em especial a validade de critério, consistindo em verificar a eficácia em prever um desempenho específico

independente do teste utilizado (podendo ser preditiva ou concorrente).

- Proposição de um instrumento multidimensional para avaliação dos estudantes da educação profissional, observando essencialmente as dimensões cognitivas e psicomotoras, e indiretamente a dimensão afetiva. Essa abordagem poderá ser interessante visto que os domínios são combinados e, somente pela sua mobilização concomitante, permitem diagnosticar a competência profissional em sua totalidade.
- Aplicação do método estabelecido em outras áreas tecnológicas, sejam na esfera industrial, comercial ou da prestação de serviços, com o objetivo de acompanhar suas especificidades e validação nas diferentes esferas. O ato comparativo permitirá ampliar o espectro de análises e pontos a serem considerados como orientação para elaboração de instrumentos dessa natureza.
- Novas abordagens de respostas de itens para mensuração do domínio psicomotor, em especial o emprego de rubricas graduais, apresentando para uma mesma evidência quatro (sugestão) possibilidades, progressivas, quanto ao comportamento esperado na tarefa. Isso permitirá comparar o caráter politômico com o dicotômico, auxiliando na definição de qual abordagem tem maior aderência para avaliação deste traço latente.

REFERÊNCIAS

ABBAD, G., et al. **Planejamento instrucional em TD&E.** In BORGES-ANDRADE, J. E.; ABBAD, G.; MOURÃO, L. (Orgs.), *Treinamento, desenvolvimento e educação em organizações e trabalho: fundamentos para a gestão de pessoas.* Porto Alegre: Artmed, 2006. p. 469-488.

ABELSON, R. P. A variance explanation paradox: When a little is a lot. **Psychological Bulletin.** 1985. vol. 97, n. 1, p. 129-133. Disponível em: <https://wolfweb.unr.edu/~ldyer/classes/396/abelson.pdf>. Acesso em: 04 abr. 2019.

ALBANESE M.T.; CAMEY S.A.; MARCONDES FILHO, D. **Análise das questões que compõem as provas do Concurso Vestibular 93-UFRGS:** usando o Modelo de Variáveis Latentes Logit-Probít. Porto Alegre: UFRGS, 1993.

ALLAL, L. **Estratégias de avaliação formativa:** concepções psicopedagógicas e modalidades de aplicação. In L. ALLAL, J.; CARDINET, P.; PERRENOUD, A. *Avaliação Formativa num Ensino Diferenciado.* Coimbra: Livraria Almedina, 1986. p. 297-342,

ANASTASI, A.; URBINA, S. **Fidedignidade:** Validade. In: *Testagem psicológica*, 7. ed. Porto Alegre: Artmed, 2000. p. 84-152.

ANDERSON, L.; KRATHWOHK, D. R.; *et al.* ***A taxonomy for learning, teaching, and assessing: a revision of Bloom's Taxonomy of Educational Objectives.*** New York: Addison, Wesley Longman, 2001.

ANDRADE D.F.; TAVARES, H.R.; VALLE, R.C. **Teoria da Resposta ao Item:** Conceitos e Aplicações. São Paulo: Associação Brasileira de Estatística, 2000.

ANDRICH, D. ***A rating formulation for ordered response categories.*** Nedlands: Psychometrika, 1978.

ARAÚJO, C. M. M.; RABELO, M. L. Avaliação Educacional: a abordagem por competências. **Avaliação Revista da Avaliação da Educação Superior (Campinas).** Campinas-SP, v. 20, n. 2, 2015. p. 443-466.

BAKER, F.B. *The basics of item response theory*. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation, 2001.

BIRNBAUM, A. *Some latent trait models and their use in inferring and examinee's ability*. In: Lord FM, Novick MR. *Statistical theories of mental test scores*. Reading, MA: Addison –Wesley, 1968.

BLOOM, B. S. *et al.* **Taxionomia de objetivos educacionais**: compêndio primeiro-domínio cognitivo. Porto Alegre: Globo, 1983.

BOCK, R. D. *Estimating item parameters and latent ability when responses are scored in two or more nominal categories*. Nedlands: Psychometrika, 1972.

BORTOLOTTI, S. L. V.; ANDRADE, D. F. Aplicação de um Modelo de Desdobramento Graduado Generalizado - GGUM da Teoria da Resposta ao Item. **Estudos em Avaliação Educacional**. São Paulo, v. 18, n 37, 2007.

BRASIL. [Constituição (1937)]. **Constituição da República Federativa do Brasil**: promulgada em 10 de novembro de 1937. Brasília: Casa Civil, 1937. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao37.htm. Acesso em: 21 jan. 2018.

BRASIL. **Lei n. 4.048, de 22 de janeiro de 1942**. Cria o Serviço Nacional de Aprendizagem dos Industriários (SENAI). Brasília: Casa Civil, 1942. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/1937-1946/Del4048.htm. Acesso em: 21 jan. 2018.

BRASIL. **Lei n. 8.621, de 10 de janeiro de 1946**. Dispõe sobre a criação do Serviço Nacional de Aprendizagem Comercial (SENAC) e dá outras providências. Brasília: Casa Civil, 1946. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/1937-1946/Del8621.htm. Acesso em: 21 jan. 2018.

BRASIL. **Lei nº 4.024, de 20 de dezembro de 1961**. Fixa as Diretrizes e Bases da Educação Nacional. Brasília: Casa Civil, 1961. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/L4024.htm. Acesso em: 22 jan. 2018.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil**. Promulgada em 5 de outubro de 1988. Brasília: Casa Civil, 1988. Disponível em: https://www2.senado.leg.br/bdsf/bitstream/handle/id/508200/CF88_EC85.pdf. Acesso em: 22 jan. 2018.

BRASIL. **Lei nº 9.394 de 20 de dezembro de 1996**. Estabelece as diretrizes e bases da educação nacional. Brasília: Casa Civil, 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/L9394.htm. Acesso em: 23 jan. 2018.

BRASIL. **Decreto-Lei nº 2.208, de 17 de abril de 1997**. Regulamenta o § 2º do art. 36 e os arts. 39 a 42 da Lei nº 9.394. Brasília: Casa Civil, 1997. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/D2208.htm. Acesso em: 23 jan. 2018.

BRASIL. **Lei n. 12.513, de 26 de outubro de 2011**. Institui o Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec). Brasília: Casa Civil, 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12513.htm. Acesso em: 24 jan. 2018.

BRASIL. **Resolução nº 6, de 20 de setembro de 2012**. Define Diretrizes Curriculares Nacionais para a Educação Profissional Técnica de Nível Médio. Brasília: MEC, 2012. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=11663-rceb006-12-pdf&category_slug=setembro-2012-pdf&Itemid=30192. Acesso em: 24 jan. 2018.

BOOKER, M. J. *A Roof without Walls: Benjamin Bloom's Taxonomy and the Misdirection of American Education*. **Academic Questions**, n. 20, p. 347-355, 2007. Disponível em: <https://rd.springer.com/article/10.1007/s12129-007-9031-9>. Acesso em: 15 fev. 2018.

BUSCHING, B. *Grading Inquiry Projects*, In: ANDERSON, R. S., SPECK, B. W. (Editors), *Changing the Way We Grade Student Performance*. São Francisco: Jossey-Bass, 1998.

CANALI, H. H. B. **A trajetória da educação profissional no Brasil e o os desafios da construção de um ensino médio integrado a educação profissional.** V Simpósio sobre trabalho e educação. Belém: Universidade Federal do Pará, 2009.

CANDAU, V. M.; LELIS, I. A.; CANDAU, V. M. (Org.). **Rumo a uma nova didática.** 15. ed. Petrópolis: Vozes, 2003.

CARBONE, P.P. **Gestão por competências e gestão do conhecimento.** Rio de Janeiro: Editora FGV, 2006.

CASTRO, S. M. J; TRENTINI, C.; RIBOLDI, J. Teoria da resposta ao item aplicada ao Inventário de Depressão Beck. **Revista Brasileira de Epidemiologia**, São Paulo, v. 13, 2010. Disponível em: <http://www.scielo.br/pdf/rbepid/v13n3/12.pdf>. Acesso em: 06 mar. 2018.

CLARKE, M.; OXMAN, A.D. (editors) **Cochrane Reviewers' Handbook 4.1 In: Review Manager (RevMan)** [Computer program]. Version 4.1. Oxford, England: The Cochrane Collaboration, 2001.

COLES C. *Developing professional judgement.* **Journal of Continuing Education in the Health Professions**, 2002. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/chp.1340220102>. Acesso em: 27 fev. 2018.

CONDE, F. N. **Análise empírica de itens.** Technical report: Instituto Nacional de Estudos e Pesquisas Educacionais - DAEB/INEP/MEC, 2001.

CONKLIN, J. *A taxonomy for learning, teaching and assessing: a revision of Blooms's taxonomy of educational objectives.* **Educational Horizons**, 2005. v. 83, n. 3, p. 153-159. Disponível em: https://www.jstor.org/stable/42926529?seq=1#page_scan_tab_contents. Acesso em: 10 fev. 2018.

CUNHA, L. A. **O Ensino de Ofícios nos Primórdios da Industrialização.** São Paulo: Unesp, Brasília: Flacso, 2000.

DAVE, R. **Psychomotor Levels**. In R. J. Armstrong (Ed.). *Developing and Writing Behavioral Objectives*. Tucson, AZ: Educational Innovators Press. 1970.

DELISLE, R. *How to use problem-based learning in the classroom*. Alexandria, Virginia, U.S.A.: Association for Supervision and Curriculum Development (ASCD), 1997.

DEMARS, C. *Item Response Theory*. New York: Oxford University, Understanding Statistics. 2010.

DEPRESBITERIS, L. **O Desafio da Avaliação da Aprendizagem: dos fundamentos a uma proposta inovadora**, São Paulo, E.P.U., 1989.

DEPRESBITERIS, L. **Avaliando Competências na Escola de Alunos ou na Escola de todos?** [Boletim Técnico do SENAC]. Rio de Janeiro: SENAC, 2001. vol. 27. n. 3.

DEPRESBITERIS, L. **Competências na educação profissional: é possível avalia-las?** [Boletim Técnico SENAC]. Rio de Janeiro: SENAC, 2005. v. 31, n.2.

DICKSON R. *Systematic Reviews In Hamer S, Collinson G. Achieving evidence-based practice: a handbook for practitioners London*. Baillière Tindall; 1999. p.41-60.

DINHAM, S.; ROEW, K. "Teaching and learning in middle schooling: a review of the literature", 2009. Disponível em: https://research.acer.edu.au/tll_misc/11. Acesso em: fev. 2018.

EDELEN, M. O.; REEVE, B. B. *Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement*. Qual Life Res, 2007. v. 16, p. 05-18. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/17375372>. Acesso em: 23 mar. 2018.

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

ESCOTT, C.M.; MORAES, M.A.C. **História da educação profissional no Brasil**: as políticas públicas e o novo cenário de formação de professores nos Institutos Federais de Educação, Ciência e Tecnologia. *In*: IX Seminário Nacional de Estudos e Pesquisas “História, Sociedade e Educação no Brasil”. João Pessoa: Universidade Federal da Paraíba, 2012. p. 1492-1508.

FAMEMA. **Necessidades de saúde**: 1ª série – cursos de medicina e enfermagem. Marília: Faculdade de Medicina de Marília, 2004.

FERRAZ, A.P.C.M.; BELHOT, R.V. **Taxonomia de Bloom**: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. São Carlos, 2010. v. 17, n. 2, p. 421-431. Disponível em: http://www.scielo.br/scielo.php?pid=S0104-530X2010000200015&script=sci_abstract&tlng=pt. Acesso em: 20 fev. 2018.

FERRAZ, M.J. *et al.* **A Avaliação Criterial /Avaliação Normativa**. I.I.E. Pensar Avaliação, Melhorar a Aprendizagem. Lisboa: I.I., 1994.

FEUERWERKER, L.C.M. **Além do discurso de mudança na educação médica**. São Paulo: Hucitec, Rede Unida, ABEM, 2002.

FIGARI, G. **Avaliar, que Referencial?** Porto: Porto Editora. 1996.

FISHER, B. T. D. **Avaliação da aprendizagem**: obsessão pelo resultado pode obscurecer a importância do processo. *In*: WERLE, F. (Org.). Avaliação em larga escala: foco na escola. São Leopoldo: Oikos; Brasília: Liber Livro, 2010. p. 21-36.

FLETCHER, P. R. A Teoria da Resposta ao Item: medidas invariantes do desempenho escolar. **Ensaio. Avaliação e Políticas Públicas em Educação**, Rio Comprido: 1994. v. 2, p. 21-28.

FLEURY, M.T.L.; FLEURY, A. Construindo o conceito de competência. RAC – **Revista de Administração Contemporânea**. ed. esp. 2001. v. 5, p. 183-196. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-6552001000500010. Acesso em: 21 jan. 2018.

FONSECA, C. S. **História do Ensino Industrial no Brasil**. Rio de Janeiro: Escola Técnica, 1961.

FREITAS, A. **A formação de professores para a educação profissional técnica de nível médio**: a experiência do Centro Paula Souza. 2010. 147 f. il. Dissertação (Mestrado em Tecnologia). São Paulo: Centro Estadual de Educação Tecnológica Paula Souza, 2010.

FREUND, J.E.; SIMON, G.A. **Estatística básica**: economia, administração e contabilidade. 9 ed. Porto Alegre: Bookman, 2000.

FURLAN, M.I.C. **Avaliação da aprendizagem escolar**: convergências, divergências. 1ª edição. São Paulo. Amablume editora, 2007.

GARCIA, S.R.O. **O fio da história: a gênese da formação profissional no Brasil**. In: Trabalho e Crítica. São Leopoldo: Ed. Unisinos, 2000.

GHIRALDELLI P. J. R. **História da educação**. 3. ed. São Paulo: Cortez, 2008.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

GOERGEN, P. L. **Teoria e prática**: problema básico da educação. In: REZENDE, A. M. (Org.) Iniciação teórica e prática às ciências da educação. Petrópolis: Vozes, 1979.

GOLDBERG, M.A.A.; SOUSA C. P. **A prática da avaliação**. São Paulo: Cortez e Moraes, 1979.

GRONLUND, N.E. *Measurement and evaluation in teaching*. 5. ed. New York: McMillan Publishing, 1985.

GUBA, E.; LINCOLN, Y. *Effective Evaluation*. San Francisco: Jossey-Bass, 1981.

HADJI, C. **Avaliação**: as Regras do Jogo. Porto: Porto Editora, 1994.

HAMBLETON, R.K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of item response theory*. Newbury Park-CA: Sage; 1991.

HAMBLETON, R. K. *Emergence of item response modeling in instrument development and data analysis*. Massachusetts: University of Massachusetts, 2000. v. 38, n. 9, p. 60-65.

HAMBLETON, R. K. *Theory, methods and practices in testing for the 21st century*. Psicothema: 2004. p. 696–701. Disponível em: <https://psycnet.apa.org/record/2004-19992-023>. Acesso em: 21 mar. 2018.

HARROW, A.J. *Taxinomia do domínio psicomotor*. Rio de Janeiro: Globo, 1983.

HILL, M. M.; HILL, A. *Investigação empírica em ciências sociais: um guia introdutório*. 2. ed. [S.l.]: Dinâmica, 1998.

HOFFMANN, J. M. L. *Avaliar para promover: as setas para o caminho*. 10. ed. Porto Alegre: Mediação, 2008.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Exame Nacional de Desempenho de Estudantes** (ENADE 2012). Relatório Síntese: Administração. 2012. Disponível em: http://download.inep.gov.br/educacao_superior/enade/relatorio_sintese/2012/2012_rel_administracao.pdf. Acesso em: 29 set. 2018.

JONASSEN, D. H. *Evaluation Constructivistic Learning, In: Constructivism and the Technology of Instruction*. Hillsdale: Lawrence Erlbaum Associates Publishers, 1992.

KLEIN, R. Utilização da Teoria da Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica – Saeb. **Meta-Avaliação**. Rio de Janeiro: 2009. v. 1, n. 2, p. 125-140. Disponível em: <http://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38>. Acesso em: 21 mar. 2018.

KLEIN, R. Alguns aspectos da teoria de resposta ao item relativos à estimação das proficiências. **Ensaio: avaliação e políticas públicas em educação**. Rio de Janeiro: 2013. v. 21, n. 78, p. 35-55. Disponível em: http://www.scielo.br/scielo.php?pid=S0104-40362013000100003&script=sci_abstract&tlng=pt. Acesso em: 22 mar. 2018.

KLOTZ, V.K.; WINTHER, E.; FESTNER, D. *Modeling the Development of Vocational Competence: a Psychometric Model for Economic Domains*. *Vocations and Learning*: 2015. p.247– 268.

KÖCHE, J. C. **Fundamentos de metodologia científica**: Teoria da ciência e iniciação à pesquisa. 20. ed. atual. Petrópolis-RJ: Vozes, 1997.

KOLEN, M. J.; BRENNAN, R. L. *Test Equating, Scaling, and Linking: methods and practices*. Iowa City, USA: Springer, 2010.

KRAEMER, M.E.P. **Avaliação da aprendizagem como construção do saber**. *V Coloquio Internacional sobre Gestión Universitaria en América del Sur*. Mar del Plata, 2005. Disponível em: <https://repositorio.ufsc.br/handle/123456789/96974>. Acesso em: 22 fev. 2018.

KRATHWOHL, D. R. *A revision of Bloom's taxonomy: an overview*. *Theory in Practice*, 2002. v. 41, n. 4, p. 212-218. Disponível em: <https://www.depauw.edu/files/resources/krathwohl.pdf>. Acesso em: 10 fev. 2018.

KUENZER, A.Z. (org). **Ensino Médio**: construindo uma proposta para os que vivem do trabalho. 5. ed. São Paulo: Cortez, 2007.

LAI, J.S.; et al. *Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item back from the FACIT-Fatigue Scale*. *Qual Life Res* 2003. p. 485-501. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/13677494>. Acesso em: 27 jan. 2018.

LE BOTERF, G. **Desenvolvendo Competências dos Profissionais**. 3. ed. Porto Alegre: Artmed, 2003.

LIBÂNEO, J.C. **Didática**. 2ª ed. São Paulo: Cortez, 1994.

LIMA, G.Z.; LINHARES, R.E.C. **Escrever bons problemas**. Revista Brasileira de Educação Médica. Rio de Janeiro: 2008. p. 197-201.
Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-55022008000200007.
Acesso em: 12 fev. 2018.

LORD, F.M. *A theory of test scores*. Psychometric Monograph, Iowa City, IA: Psychometric Society, 1952.

LORD, F.M. *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: LEA, 1980.

LUCKESI, C. C. **Avaliação da aprendizagem escolar**: estudos e proposições. São Paulo: Cortez, 2011.

LUDKE, M. **O Trabalho com Projetos e a Avaliação na Educação Básica**. In: ESTEBAN, M.T.; HOFFMANN, J.; SILVA, J.F. (orgs) Práticas Avaliativas e Aprendizagens Significativas. Porto Alegre: Mediação, 2004. p.73-76.

MACEDO, L. **Situação-problema**: Forma e recurso de avaliação, desenvolvimento de competências e aprendizagem escolar. In: PERRENOUD, P; et al. (Org.). As competências para ensinar no século XXI: A formação dos professores e o desafio da avaliação. Porto Alegre: Artmed Editora, 2002. p. 113-135.

MACEDO, L. **Competências e habilidades**: elementos para uma reflexão pedagógica e Propostas para pensar sobre situações-problema. In: BRASIL. Ministério da Educação. INEP. Exame Nacional do Ensino Médio (Enem): fundamentação teórico-metodológica. Brasília, 2005. p.13-28 e 37-39.

MAGER, R. **A formulação de objetivos de ensino**. Porto Alegre: Editora Globo, 1979.

MANACORDA, M. A. **História da educação**: da Antiguidade aos nossos dias. 4. ed. São Paulo: Cortez, 1995.

MANFREDI, S.M. **Educação profissional no Brasil**. São Paulo: Cortez, 2002.

MARCONI, M. A.; LAKATOS, E. M. **Técnicas de pesquisa**. 3. ed. São Paulo: Atlas, 1999.

MARTINS, G. A. THEÓPHILO, C. R. **Metodologia da investigação científica para ciências sociais aplicadas**. 2.ed. São Paulo: Atlas, 2009.

MASTERS, G.N. *A Rasch model for partial credit scoring*. Psychometrika: 1982. p. 149-174. Disponível em: <https://rd.springer.com/article/10.1007/BF02296272>. Acesso em: 19 mar. 2018.

MATTHIENSEN, A. **Uso do Coeficiente Alfa de Cronbach em Avaliações por Questionários**. Boa Vista: [s.n.], 2011.

MEDRONHO, R. A.; *et al.* **Epidemiologia**. São Paulo: Editora Atheneu, 2003.

MEIRIEU, P. **Aprender sim, mas como?** 7. ed. Porto Alegre: Editora Artmed, 1998.

MILLER, G.E. *The assessment of clinical skills/competence/performance*. Academic Medicine, 1990. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/2400509>. Acesso em: 15 fev.

MILONE, G. **Estatística Geral e Aplicada**. São Paulo: Pioneira Thomson Learning, 2004.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 2. ed. Rio de Janeiro: Editora LTC, 2003.

MORRETO, V.P. **Planejamento: planejando a educação para o desenvolvimento de competências**. 5. ed. Petrópolis: Vozes, 2010.

MUÑIZ, J. *Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa*. Madri: Ediciones Pirámide, S.A., 1990.

MURAKI, E. *A generalized partial credit model: Application of an EM algorithm*. Applied Psychological Measurement, 1992. v. 16, p. 159–176. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2333-8504.1992.tb01436.x>. Acesso em: 25 mar. 2018.

NERY, A. **Parecer sobre a Matriz Curricular de Língua Portuguesa**. 3. ed. Mimeo Brasília: 2000.

NICHOLSON, P. *et al.* **Measuring nursing competencies in the operating theatre**: Instrument development and psychometric analysis using Item Response Theory. Nurse Education Today, 2013. p. 1088–1093. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/22608826>. Acesso em: 20 jan. 2018.

NORCINI, J. *et al.* **Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference**. Med Teach, 2011. v. 33, n. 3, p.206 – 214. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/21345060>. Acesso em: 25 fev. 2018.

NUNES, C. H. S. da S.; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por Teoria de Resposta ao Item. **Avaliação Psicológica**, 2005. v. 4, n. 2, p. 141-153. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712005000200006. Acesso em: 04 abr. 2018.

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

O'GRADY, K. E. Measures of explained variance: Cautions and limitations. **Psychological Bulletin**. 1982. vol. 92, n. 3, p. 766-777. Disponível em: <https://psycnet.apa.org/record/1983-04833-001>. Acesso em: 04 abr. 2019.

OBSERVATÓRIO DO PNE. **Educação Profissional**. Disponível em: <http://www.observatoriodopne.org.br/metas-pne/11-educacao-profissional>. Acesso em: 21 jan. 2018.

ORLANDO, M.; MARSHALL, G. N. *Differential item functioning in a Spanish translation of the PTSD checklist: Detection and evaluation of impact*. Psychological Assessment, 2002. v. 14, n. 1, p.50–59.

PASQUALI, L. **Psicometria: teoria e aplicações**. Brasília: Editora Universidade de Brasília, 1997.

PASQUALI, L. **Instrumentos psicológicos: manual prático de elaboração**. Brasília: LabPAM/ IBAPP, 1999.

PASQUALI, L.; PRIMI, R. Fundamentos da Teoria da resposta ao Item. **Avaliação Psicológica**. Porto Alegre, 2003. v. 2, n. 2, p. 99-110.

PASQUALI L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes; 2004.

PASQUALI, L. Psicometria. **Revista da Escola de Enfermagem**. São Paulo: USP, 2009. v. 43, p. 992-999.

PAZIN FILHO, A.; SCARPELINI, S. Simulação: definição. *In*: Simpósio Didática II – Simulação. **Revista Medicina**. Ribeirão Preto, 2007. v. 40, n. 2, p. 162-166.

PENIN, S.; MARTÍNEZ, M. **Profissão docente: pontos e contrapontos**. São Paulo: Summus, 2009.

PERRENOUD, P. **Construir as competências desde a escola**. Porto Alegre: Artmed Editora; 1999.

PERRENOUD, P. **Dez novas competências para ensinar**. Porto Alegre: Artmed, 2000.

POPHAN, W.J. *Educacional evaiuation*. Engleword Cliffs: Allyn & Bacon, 1975.

PRIMI, R.; ALMEIDA, L. S. **Considerações sobre a análise factorial de itens com resposta dicotômica**. Psicologia: Teoria, Investigação e Prática, 1998. p. 225-234.

RASCH, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research, 1960.

RICHARDSON, R. J. **Pesquisa Social: Métodos e Técnicas**. Colaboradores José Augusto de Souza Peres. São Paulo: Editora Atlas, 1989.

SAMEJINA, F. *Estimation of latent ability using a response pattern of graded scores*. Psychometrika Monograph Supplement, 1969. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2333-8504.1968.tb00153.x>. Acesso em: 18 mar. 2018.

SANT'ANNA, I.M. **Por que avaliar? Como avaliar?** Critérios e instrumentos. 3. ed. Petrópolis-RJ: Vozes, 1995.

SANTOS, C. R., et al. **Avaliação Educacional: um olhar reflexivo sobre sua prática**. São Paulo: Editora Avercamp, 2005.

SANTOS, J.A. **A trajetória da educação profissional**. In: LOPES, et al. (org.). 500 anos de educação no Brasil. 2. ed. Belo Horizonte: Autêntica, 2000.

SANTOS, L. M. **Desempenho escolar em Pernambuco: análise dos itens e das habilidades usando Teoria Clássica e TRI**. Dissertação (Mestrado em Estatística). Recife: Universidade Federal de Pernambuco, 2008.

SAVIANI, D. **História das ideias pedagógicas no Brasil**. Campinas-SP: Autores Associados, 2007.

SCHMITT, N. *Uses and abuses of coefficient alpha*. **Psychological Assessment**: 1996. v.8, p.350-353. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.617.5270&rep=rep1&type=pdf>. Acesso em: 30 ago. 2018.

SCHNICK-VOLLMER, K. S.; *et al.* *Modeling the Competencies of Prospective Business and Economics Teachers: Professional Knowledge in Accounting*. **Zeitschrift für Psychologie**. 2015. v. 223, p. 24-30.

SEIDEL, T.; STÜRMER, K. *Modeling and Measuring the Structure of Professional Vision in Preservice Teachers*. **American Educational Research Journal**: 2014. v. 51, n. 4, p. 739-771. Disponível em: <https://psycnet.apa.org/record/2014-29967-004>. Acesso em: 25 jan. 2018.

SENAI. Departamento Nacional. **Metodologia SENAI de Educação Profissional**. Brasília: SENAI/DN, 2013.

SHEN, L. *Progress Testing for Postgraduate Medical Education: a Four-Year Experiment of American College of Osteopathic Surgeons Resident Examinations*. **Advances in Health Sciences Education** 5: 117–129, 2000.

SILVA, A. C. R. **Metodologia da pesquisa aplicada à contabilidade**. 3. ed. São Paulo: Atlas, 2010.

SILVA, J. F. **Avaliação na Perspectiva Formativa-Reguladora: pressupostos teóricos e práticos**. Porto Alegre: Mediação, 2004.

SIMPSON, E. **The classification of educational objectives in the psychomotor domain**: The psychomotor domain. Washington, DC: Gryphon House; 1972.

SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa Operacional**. 2005. v. 25, n. 1, p. 83-112.

SOUZA, S. Z. L.; OLIVEIRA, R. P. Políticas de avaliação da educação e quase mercado no Brasil. **Educação e Sociedade**, Campinas: 2003. v. 24, n. 84, p. 873-895. Disponível em: <http://www.scielo.br/pdf/es/v24n84/a07v2484.pdf>. Acesso em: 18 fev. 2018.

SPEARMAN C. *"General Intelligence", objectively determined and measured*. Am J Psychol. 1904.

STERNBERG, R. J. **Psicologia cognitiva**. Porto Alegre: Artmed, 2000.

TABACHNICK, B.; FIDELL, L. S. **Using multivariate statistics**. 3. ed. New York: Harper Collins, 1996.

TAGGART, G.L. *et al.* **Rubricas: a handbook for construction and use**. Lanham-NY: Rowman & Littlefield Education, 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. Salvador-BA: 2009. Disponível em: http://www.abepro.org.br/biblioteca/enegep2009_TN_STP_094_638_13156.pdf. Acesso em: 04 mar. 2018.

THURSTONE LL. *A Law of comparative judgment*. **Psychol Transactions**. 1994. v. 7, n. 4, p.328. Disponível em: <https://psycnet.apa.org/record/1928-00527-001>. Acesso em: 05 abr. 2018.

TURRA, C. M. G.; *et al.* **Planejamento de ensino e avaliação**. Porto Alegre: Sagra, 1998.

VALENTINI, F.; HAUCK FILHO, N. Estabilidade da estimação dos parâmetros dos itens em diferentes tamanhos amostrais. **III CONBATRI**. 2013. n. 1, p. 43-48. Disponível em: <http://abave.com.br/ojs/index.php/conbratri/article/view/78>. Acesso em: 08 abr. 2018.

VANCE, S.; BOSWORTH, A. **Replicating reality: the power of simulations**. Salt Lake City, 2003. Disponível em: <http://www.allencomm.com/webminars/moreinfo.aspx? CourseID=8>. Acesso em: 11 abr. 2018.

VARGAS, V. do C. C. de. **Medida padronizada para avaliação de intangíveis organizacionais por meio da teoria da resposta ao item**. Tese Doutorado (Programa de Pós-Graduação em Engenharia de Produção). Florianópolis: Universidade Federal de Santa Catarina, 2007.

VENDRAMINI, C. M. M.; DIAS, A. S. **Teoria de resposta ao item na análise de uma prova de estatística em universitários**. Psico-USF (impr.) Itatiba, 2005. v. 10, n. 2, p. 201-210, Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sciarttext&pid=S1413-82712005000200012>. Acesso em: 22 jan. 2019.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. São Paulo: Atlas, 2013.

VIANNA, Heraldo Marelim. **Fundamentos de um programa de avaliação educacional**. Brasília: Liber Livro, 2005.

ZANELLA, L. **Metodologia de Estudo e de Pesquisa em Administração**. Brasília: CAPES, 2009.

ZIMOWSKI, M. F., et al.; ***BILOG-MG: Multiplegroup IRT analysis and test maintenance for binary items***. Chicago: Scientific Software International, 1996.

APÊNDICE A – Script R: dimensionalidade do teste

Script inserido no R para gerar gráfico *scree plots* e cálculo das cargas fatoriais em relação a dimensionalidade do teste aplicado aos estudantes do Curso Técnico em Mecânica.

```
# Carregar arquivo saep
apply(saep, 2, table)

# Análise de componentes principais com análise paralela
library(psych)
fa.parallel(saep, cor="tet", fa="pc", sim=FALSE)

#Análise fatorial
library(mirt)
m1 <- mirt(saep, 1, itemtype = '2PL')
print(m1)
coef(m1, IRTpars = TRUE, simplify=TRUE)
summary(m1)
m2 <- mirt(saep, 2, itemtype = '2PL')
print(m2)
coef(m2, simplify=TRUE)
summary(m2)
```

APÊNDICE B – Sintaxe Bilog-MG

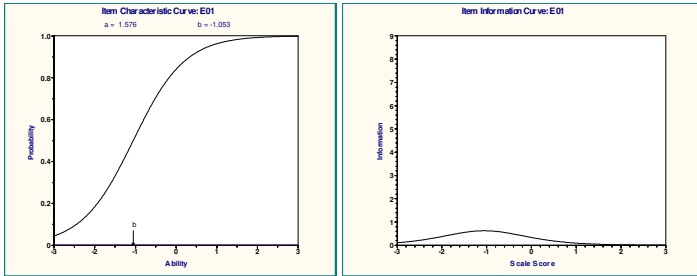
Sintaxe inserida no Bilog-MG para calibração das evidências e estimação das proficiências do domínio psicomotor, relativa aos estudantes avaliados no Curso Técnico em Mecânica.

Avaliação Domínio Psicomotor	
Curso Técnico em Mecânica	
>GLOBAL	NPArm=2,LOGistic, DFName='mecanica.DAT', SAVE;
>SAVE	PARm='mecanica.par', SCORE='mecanica.sco';
>LENGTH	NITems=32;
>INPUT	NTotal=32,NALt=2,NIDch=5, KFName='mecanica.DAT';
>ITEMS	INUMBERS=(1(1)32),INAMES=(E01(1)E32);
>TEST1 (5A1,1X,32A1)	TNAme='mecanica',INUMBER=(1(1)32);
>CALIB	NQPT=20,CYCles=50,Newton=0,CRIIt=0.001;
>SCORE	METHOD=2, NOPRINT;

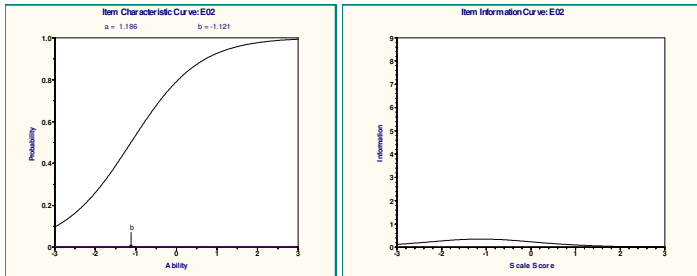
APÊNDICE C - Curvas características das evidências e curvas de informação das evidências

O gráfico do lado esquerdo representa a CCI, enquanto o da direita refere-se a CII. Para fins deste estudo, entende-se item como evidência.

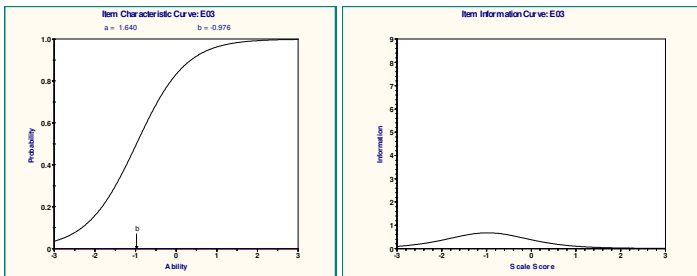
Evidência 01 – E01



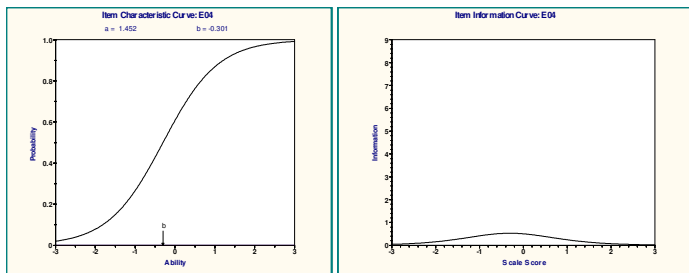
Evidência 02 – E02



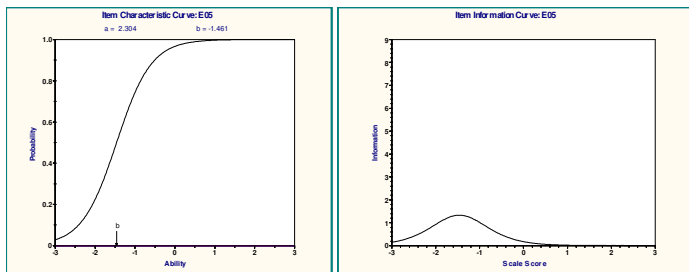
Evidência 03 – E03



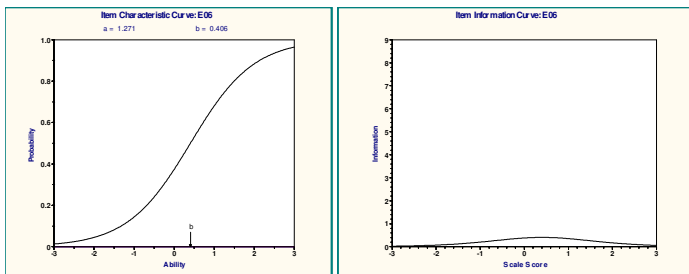
Evidência 04 – E04



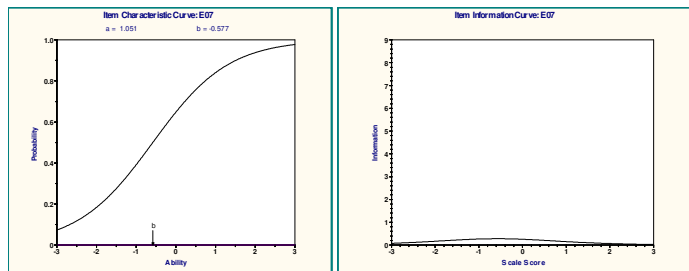
Evidência 05 – E05



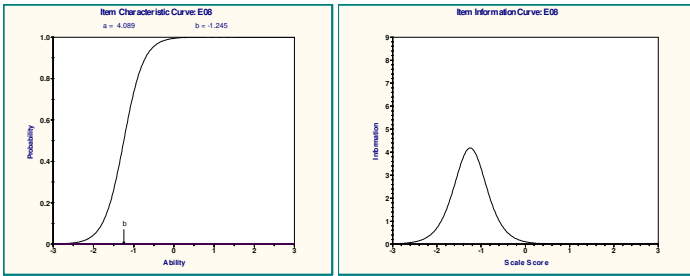
Evidência 06 – E06



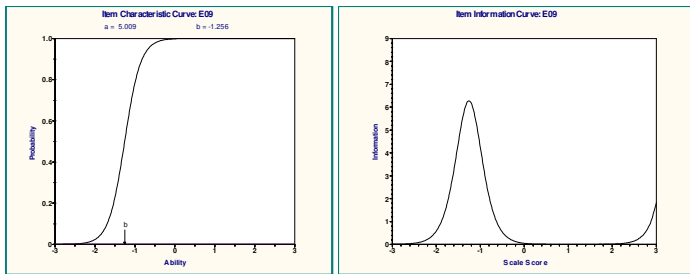
Evidência 07 – E07



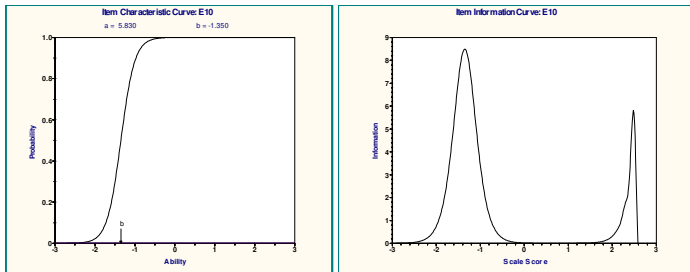
Evidência 08 – E08



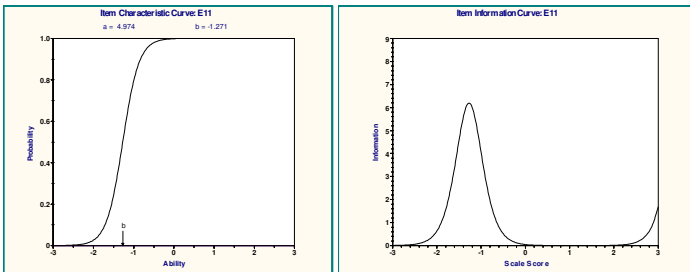
Evidência 09 – E09



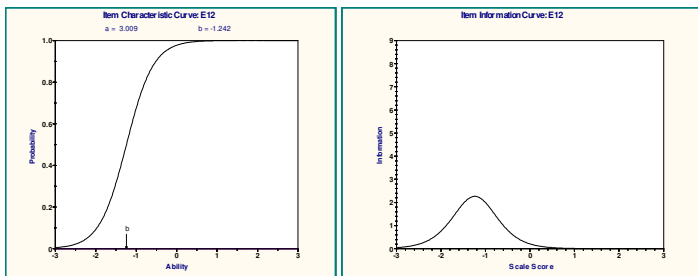
Evidência 10 – E10



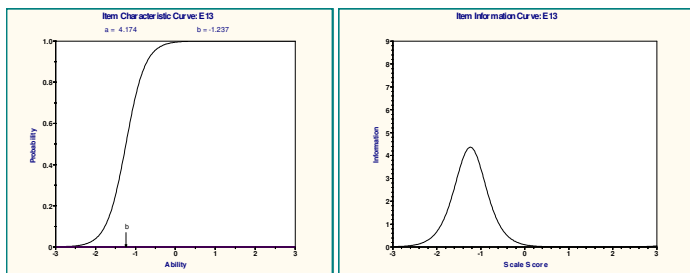
Evidência 11 – E11



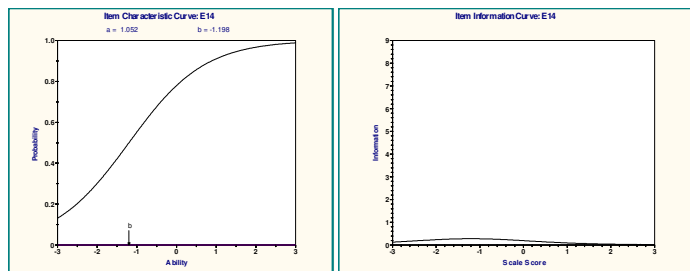
Evidência 12 – E12



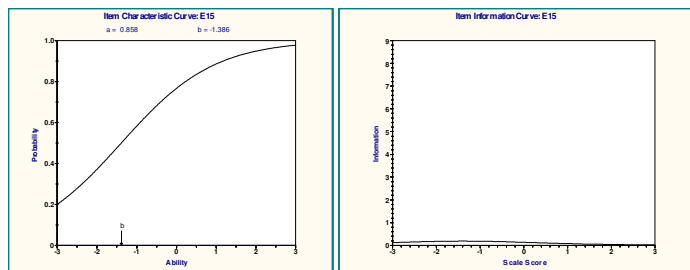
Evidência 13 – E13



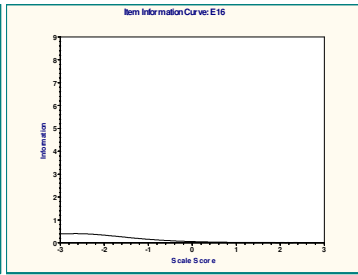
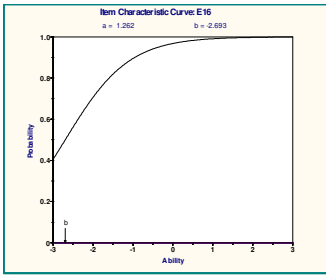
Evidência 14 – E14



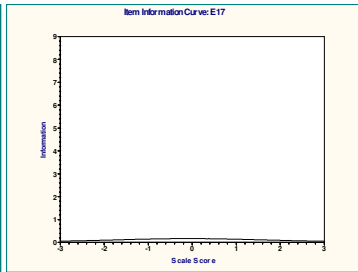
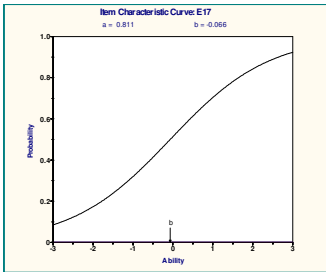
Evidência 15 – E15



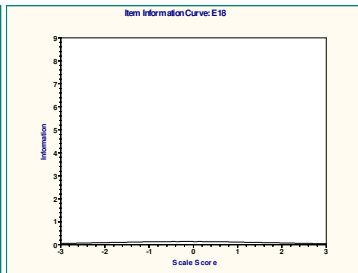
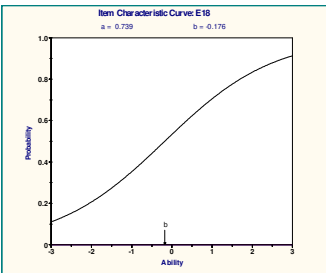
Evidência 16 – E16



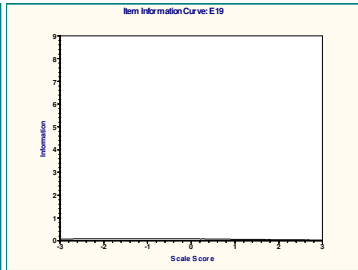
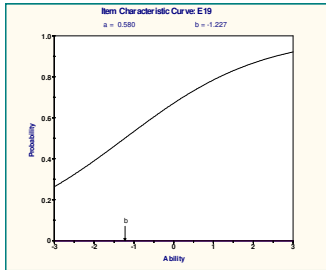
Evidência 17 – E17



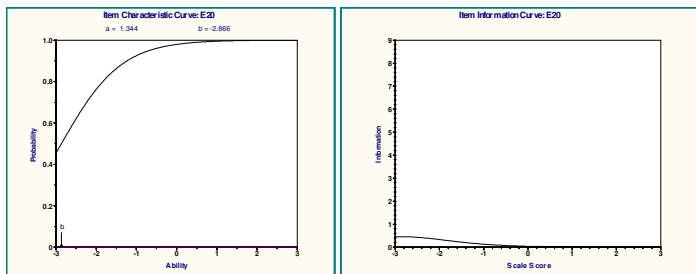
Evidência 18 – E18



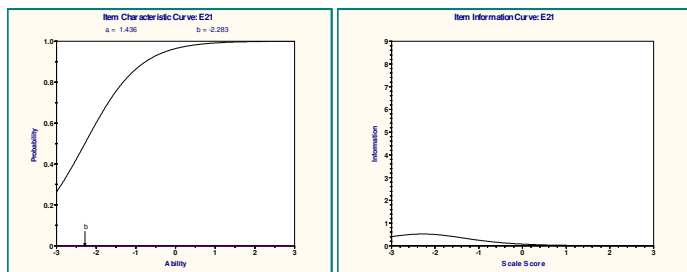
Evidência 19 – E19



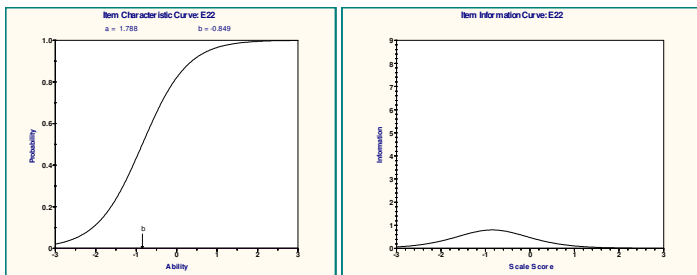
Evidência 20 – E20



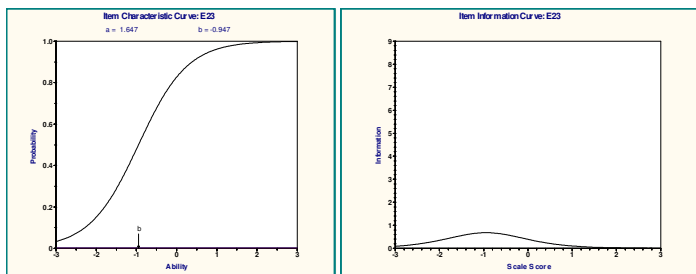
Evidência 21 – E21



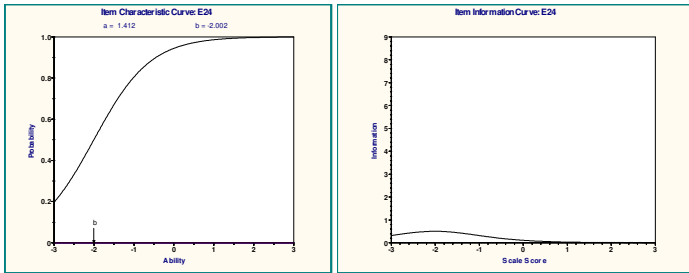
Evidência 22 – E22



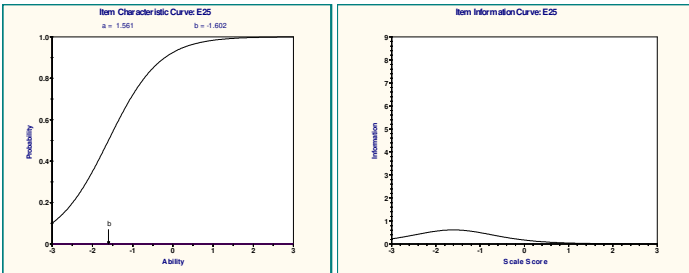
Evidência 23 – E23



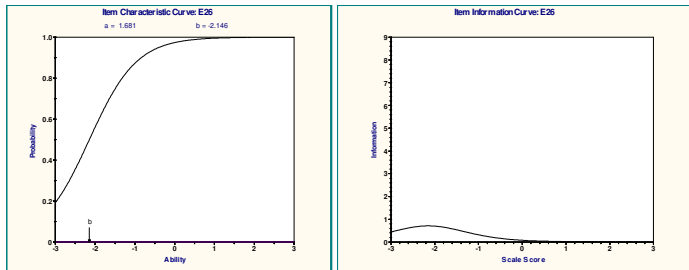
Evidência 24 – E24



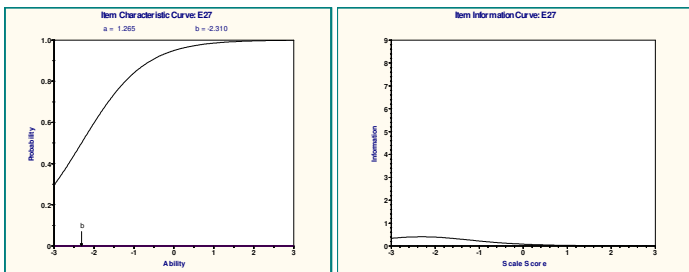
Evidência 25 – E25



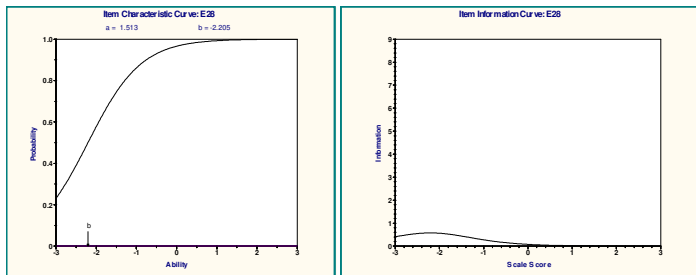
Evidência 26 – E26



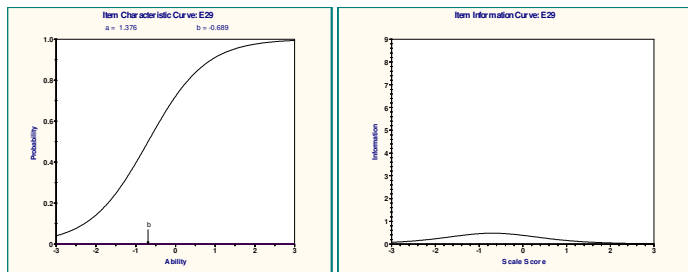
Evidência 27 – E27



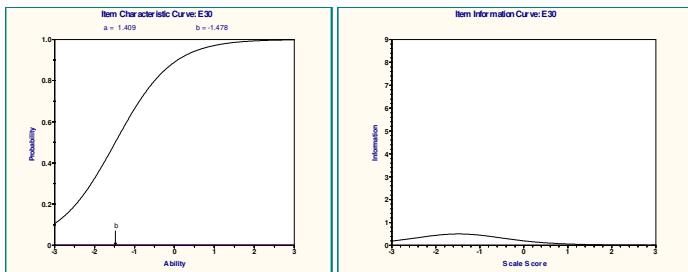
Evidência 28 – E28



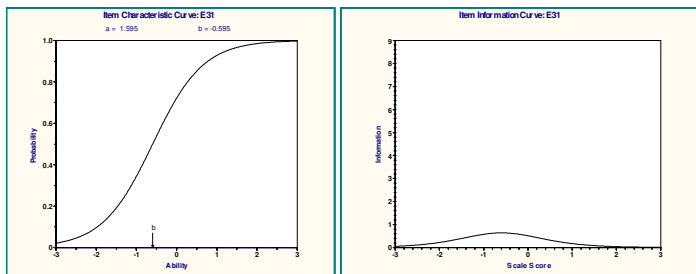
Evidência 29 – E29



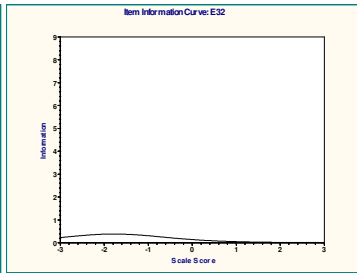
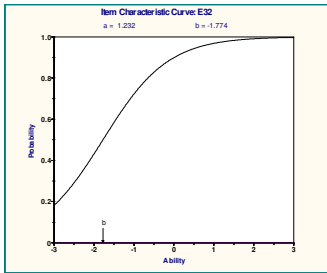
Evidência 30 – E30



Evidência 31 – E31



Evidência 32 – E32



APÊNDICE D - Posicionamento das evidências na escala de proficiência

Item	Parâmetros		Níveis da Escala																
	a	b	-4	-3,5	-3	-2,5	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3	3,5	4
			60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140
1	1,576	-1,053	0,01	0,02	0,04	0,09	0,18	0,33	0,52	0,71	0,84	0,92	0,96	0,98	0,99	1,00	1,00	1,00	1,00
2	1,186	-1,121	0,03	0,06	0,10	0,16	0,26	0,39	0,54	0,68	0,79	0,87	0,93	0,96	0,98	0,99	0,99	1,00	1,00
3	1,640	-0,976	0,01	0,02	0,03	0,08	0,16	0,30	0,49	0,69	0,83	0,92	0,96	0,98	0,99	1,00	1,00	1,00	1,00
4	1,452	-0,301	0,00	0,01	0,02	0,04	0,08	0,15	0,27	0,43	0,61	0,76	0,87	0,93	0,97	0,98	0,99	1,00	1,00
5	2,304	-1,461	0,00	0,01	0,03	0,08	0,22	0,48	0,74	0,90	0,97	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
6	1,271	0,406	0,00	0,01	0,01	0,02	0,04	0,08	0,14	0,24	0,37	0,53	0,68	0,80	0,88	0,93	0,96	0,98	0,99
7	1,051	-0,577	0,03	0,04	0,07	0,12	0,18	0,27	0,39	0,52	0,65	0,76	0,84	0,90	0,94	0,96	0,98	0,99	0,99
8	4,089	-1,245	0,00	0,00	0,00	0,01	0,04	0,26	0,73	0,95	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
9	5,009	-1,256	0,00	0,00	0,00	0,00	0,02	0,23	0,78	0,98	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
10	5,830	-1,350	0,00	0,00	0,00	0,00	0,02	0,29	0,88	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
11	4,974	-1,271	0,00	0,00	0,00	0,00	0,03	0,24	0,79	0,98	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
12	3,009	-1,242	0,00	0,00	0,01	0,02	0,09	0,32	0,67	0,90	0,98	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
13	4,174	-1,237	0,00	0,00	0,00	0,01	0,04	0,25	0,73	0,96	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
14	1,052	-1,198	0,05	0,08	0,13	0,20	0,30	0,42	0,55	0,68	0,78	0,86	0,91	0,94	0,97	0,98	0,99	0,99	1,00
15	0,858	-1,386	0,10	0,14	0,20	0,28	0,37	0,48	0,58	0,68	0,77	0,83	0,89	0,92	0,95	0,97	0,98	0,99	0,99
16	1,262	-2,693	0,16	0,27	0,40	0,56	0,71	0,82	0,89	0,94	0,97	0,98	0,99	0,99	1,00	1,00	1,00	1,00	1,00
17	0,811	-0,066	0,04	0,06	0,08	0,12	0,17	0,24	0,32	0,41	0,51	0,61	0,70	0,78	0,84	0,89	0,92	0,95	0,96
18	0,739	-0,176	0,06	0,08	0,11	0,15	0,21	0,27	0,35	0,44	0,53	0,62	0,70	0,78	0,83	0,88	0,91	0,94	0,96
20	1,344	-2,866	0,18	0,30	0,46	0,62	0,76	0,86	0,92	0,96	0,98	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00
21	1,436	-2,283	0,08	0,15	0,26	0,42	0,60	0,75	0,86	0,93	0,96	0,98	0,99	1,00	1,00	1,00	1,00	1,00	1,00
22	1,788	-0,849	0,00	0,01	0,02	0,05	0,11	0,24	0,43	0,65	0,82	0,92	0,96	0,99	0,99	1,00	1,00	1,00	1,00
23	1,647	-0,947	0,01	0,01	0,03	0,07	0,15	0,29	0,48	0,68	0,83	0,92	0,96	0,98	0,99	1,00	1,00	1,00	1,00
24	1,412	-2,002	0,06	0,11	0,20	0,33	0,50	0,67	0,80	0,89	0,94	0,97	0,99	0,99	1,00	1,00	1,00	1,00	1,00
25	1,561	-1,602	0,02	0,05	0,10	0,20	0,35	0,54	0,72	0,85	0,92	0,96	0,98	0,99	1,00	1,00	1,00	1,00	1,00
26	1,681	-2,146	0,04	0,09	0,19	0,36	0,56	0,75	0,87	0,94	0,97	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00
27	1,265	-2,310	0,11	0,18	0,29	0,44	0,60	0,74	0,84	0,91	0,95	0,97	0,99	0,99	1,00	1,00	1,00	1,00	1,00
28	1,513	-2,205	0,06	0,12	0,23	0,39	0,58	0,74	0,86	0,93	0,97	0,98	0,99	1,00	1,00	1,00	1,00	1,00	1,00
29	1,376	-0,689	0,01	0,02	0,04	0,08	0,14	0,25	0,39	0,56	0,72	0,84	0,91	0,95	0,98	0,99	0,99	1,00	1,00
30	1,409	-1,478	0,03	0,05	0,10	0,19	0,32	0,49	0,66	0,80	0,89	0,94	0,97	0,99	0,99	1,00	1,00	1,00	1,00
31	1,595	-0,595	0,00	0,01	0,02	0,05	0,10	0,19	0,34	0,54	0,72	0,85	0,93	0,97	0,98	0,99	1,00	1,00	1,00
32	1,232	-1,774	0,06	0,11	0,18	0,29	0,43	0,58	0,72	0,83	0,90	0,94	0,97	0,98	0,99	0,99	1,00	1,00	1,00

ANEXO A – Caderno de prova

O caderno de prova é composto por capa, situação-problema (desafio) e os anexos com informações relevantes para resolução (como desenhos, tabelas de referência e *templates*). Na imagem abaixo recorte apenas da contextualização e detalhamento das atividades que o estudante deveria realizar.

SITUAÇÃO-PROBLEMA - DESAFIO

A empresa Matrix Ltda. é uma empresa fabricante de componentes e acessórios para implementos agrícolas e atende diferentes regiões do Brasil.

Além da fabricação de produtos próprios, a empresa presta serviços de manutenção e reparo para clientes localizados nas regiões próximas.

Para a fabricação de seus produtos e prestação de serviços de manutenção, a empresa mantém setores como o de projetos, fabricação, manutenção, tratamento térmico, acabamento, embalagens, entre muitos outros.

Um dos fatores críticos de sucesso da empresa, que a torna reconhecida no mercado de peças para implementos agrícolas, é a existência de um processo produtivo organizado, sistematizado e controlado por meio de um sistema de gestão de qualidade.

Um dos clientes regionais encaminhou para análise técnica o eixo MTX004, que indicou a necessidade de se realizar uma usinagem e montagem de um rolamento. O orçamento encaminhado pelo departamento técnico da empresa foi aprovado pelo cliente.

Você, Técnico em Mecânica da empresa, recebeu a Ordem de Serviço do eixo MTX004 e o croqui encaminhados pelo Departamento Técnico.

A partir da ordem de serviço recebida, você terá que seguir as seguintes etapas básicas do processo técnico definidas pela empresa, que são:

- elaborar o desenho técnico completo com auxílio de software, a partir do croqui;
- preencher a ordem de serviço;
- usinar a peça conforme o desenho;
- identificar se o diâmetro efetivo do rebaixo onde será montado o rolamento é compatível com o valor estabelecido no catálogo técnico para a montagem de um rolamento submetido à carga rotativa leve no anel interno, registrando as informações sobre o eixo e o rolamento na ordem de serviço;
- realizar a montagem do rolamento no eixo.

Estará à sua disposição o croqui da peça, ordem de serviço, um microcomputador com software para elaborar desenho técnico, informações técnicas de rolamentos, tabela de ajustes do sistema ISO, torno mecânico, ferramentas para usinagem, instrumentos de medição, rolamento, dispositivo para montagem de rolamentos, tabela para parâmetros de corte.

Ao realizar as atividades, lembre-se de manter atenção às normas de saúde e segurança no trabalho e ambientais.

Ao término da atividade você deverá entregar ao departamento técnico responsável:

- desenho técnico em meio eletrônico;
- ordem de serviço preenchida e encerrada;
- eixo com o rolamento montado.

O prazo TOTAL de execução é de 3h.

ANEXO B – Lista de infraestrutura

Sistema de Avaliação da Educação Profissional Avaliação Prática de Desempenho dos Estudantes			
Curso Técnico em		Mecânica	
INFRAESTRUTURA NECESSÁRIA PARA MONTAGEM DE PROVAS POR UNIDADE			
RECURSOS MATERIAIS E EQUIPAMENTOS			
Nº	Equipamentos em geral	unid	Q. Unitária
1	Computador com software de desenho assistido por computador.	unid	1
2	Impressora A4	unid	1
3	Calculadora	unid	1
4	Bancada de trabalho (desmontagem e montagem do redutor)	unid	1
5	Torno mecânico e acessórios para usinagem	unid	1
Instrumentos		unid	Q. Unitária
1	Paquímetro universal 150 mm	unid	1
2	Rugosímetro	unid	1
3	Micrômetro interno 35 - 40 centesimal	unid	1
4	Subito 35/100mm (opcional em substituição ao Micrômetro Interno 35/40)	unid	1
Ferramentas		unid	Q. Unitária
1	Pincel 2" para limpeza	unid	1
2	Kit chaves mm e/ou pol. (conforme redutor disponível da unidade)	unid	1
3	Jogo de vazadores (conforme modelo de redutor da unidade)	unid	1
4	Martelo macete (polímero ou anti-retrocesso)	unid	1
5	Martelo de bola	unid	1
...	...	unid	1
INSUMOS NECESSÁRIOS PARA EXECUÇÃO DE PROVA POR ALUNO			
RECURSOS MATERIAIS E EQUIPAMENTOS			
Nº	Material de consumo	unid	Q. Unitária
1	Aço 1010/1020 redondo 130 x 25,4 mm	unid	1
2	Junta de papel (espessura conforme o modelo de redutor da unidade)	m	1 para 15 Aluno
3	Pano para limpeza	unid	2
Equipamentos de Proteção Individual - EPI			
Nº	EPI	Unid	Q. Unitária
1	Creme protetivo para mãos (luva química)	unid	1
2	Calçado de segurança	par	1
3	Protetor auricular (opcional - conforme procedimento da unidade)	unid	1
4	Óculos de segurança	unid	1

ANEXO C – Manual do avaliador

O manual do avaliador contempla as informações de instruções para preparação, aplicação e correção da prova (gabaritos quando assim necessários).

Instruções para preparação, aplicação e correção da prova

O Manual do Avaliador é o 'manual' de preparação, aplicação e correção das provas, de uso e acesso exclusivo do avaliador. Contém informações complementares ao Caderno de Prova do Estudante e que NÃO DEVEM ser do conhecimento do aluno. Deve ser incluída toda a informação necessária para que o avaliador, no momento de aplicar a prova, não seja obrigado a improvisar qualquer aspecto por falta de definições prévias e para que as sucessivas aplicações da prova, ainda que realizadas por distintos avaliadores e em diferentes locais, possam ser reproduzidas nas mesmas condições.

Este processo avaliativo incide sobre o curso Técnico de Mecânica, precisamente nas Unidades de Competência e seus desdobramentos em Elementos de Competência (indicadores), Padrões de Desempenho (critérios de avaliação), Capacidades Básicas, Técnicas e de Gestão, bem como nos Objetos de Conhecimento.

Os desempenhos observáveis estão explicitados por meio de evidências estabelecidas na Lista de Verificação. Essa Lista de Verificação deverá ser preenchida, assinalando com 'X' na coluna SIM ou NÃO da Escala, conforme o alcance ou não do desempenho por parte do aluno.

Leia atentamente o Caderno de Prova do Estudante e a Lista de Verificação, identificando a situação-problema e as evidências observáveis para a correta aplicação da prova. Caso tenha alguma dúvida quanto às situações-problema, referente a qualquer um dos itens da Lista de verificação, para a preparação do ambiente da prova prática ou para a aplicação ou correção da prova, você deve esclarecê-las antes de sua aplicação.

O avaliador deverá orientar os alunos quanto à administração do tempo TOTAL de 3h para cumprir as três Unidades de Competência e ambientação de 30 minutos divididos em 10 minutos para cada Unidade de Competência, totalizando 3h e 30 minutos por aluno.

Para cada Unidade de Competência foi estimado o tempo de uma hora. Caso o aluno ultrapasse esse limite, o avaliador poderá interromper e encaminhá-lo para a etapa seguinte, proporcionando ao aluno a condição de completar o processo avaliativo. Caso o aluno termine a Unidade de Competência antes de uma hora, encaminhar o aluno para a próxima etapa sem parada de tempo.

Cada avaliador poderá ser responsável pela avaliação de no máximo 4 alunos por vez.

ANEXO D – Lista de verificação

Ev.	EVIDÊNCIA	ATIVIDADE (ENTREGA)	CAPACIDADE	OBJETOS DE CONHECIMENTO
E1	Apresentou o número de vistas conforme o croqui apresentado, anexo 1	desenho técnico em meio eletrônico	C1: Interpretar dados e informações	Desenho assistido por computador; Interpretação de desenho técnico
E2	Utilizou os tipos de linhas conforme aplicação	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico; Normas técnicas e legislações
E3	Representou no desenho todas as cotas do croqui, conforme o anexo 1. (mínimo 6 de comprimento e 5 de diâmetro)	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico; Normas técnicas e legislações
E4	Representou no desenho o furo roscado conforme anexo 1	desenho técnico em meio eletrônico	C4: Definir técnicas	Desenho assistido por computador; Interpretação de desenho técnico; Normas técnicas e legislações
E5	Representou no desenho o canal conforme anexo 1	desenho técnico em meio eletrônico	C1: Interpretar dados e informações	Desenho assistido por computador; Interpretação de desenho técnico
E6	Representou no desenho os furos de centro conforme anexo 1	desenho técnico em meio eletrônico	C4: Definir técnicas	Desenho assistido por computador; Interpretação de desenho técnico; Normas técnicas e legislações

CODIGO	EVIDÊNCIA	ATIVIDADE (ENTREGA)	CAPACIDADE	OBJETOS DE CONHECIMENTO
E7	Preencheu a legenda (Nome do Desenhista, Escala, escola e Material)	desenho técnico em meio eletrônico	C1: Interpretar dados e informações	Desenho assistido por computador
E8	Desenhou o rebaixo A conforme anexo 1	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico
E9	Desenhou o rebaixo B conforme anexo 1	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico
E10	Desenhou o corpo central conforme anexo 1	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico
E11	Desenhou o rebaixo C conforme anexo 1	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico
E12	Desenhou o rebaixo D conforme anexo 1	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico

CODIGO	EVIDÊNCIA	ATIVIDADE (ENTREGA)	CAPACIDADE	OBJETOS DE CONHECIMENTO
E13	Desenhou o rebaixo E conforme anexo 1	desenho técnico em meio eletrônico	C8: Aplicar procedimentos e normas técnicas	Desenho assistido por computador; Interpretação de desenho técnico
E14	Registrou os parâmetros de corte conforme o material e ferramenta especificados na Ordem de Serviço	ordem de serviço preenchida e encerrada	C1: Interpretar dados e informações	Cálculo técnico aplicado; Tecnologia de ferramentas de corte; Tecnologia de materiais
E15	Descreveu a sequência de usinagem na Ordem de Serviço	ordem de serviço preenchida e encerrada	C2: Planejar etapas do processo	Interpretação de desenho técnico; Processos de fabricação
E16	Utilizou instrumento de medição conforme especificação técnica do projeto	eixo com o rolamento montado	C3: Especificar recursos	Interpretação de desenho técnico; Metrologia
E17	Usinou o diâmetro do rebaixo D conforme anexo 3	eixo com o rolamento montado	C5: Controlar processo	Interpretação de desenho técnico; Processos de fabricação; Tecnologia de ferramentas de corte

CODIGO	EVIDÊNCIA	ATIVIDADE (ENTREGA)	CAPACIDADE	OBJETOS DE CONHECIMENTO
E18	Usinou o rebaixo D com o comprimento especificado conforme anexo 3	eixo com o rolamento montado	C5: Controlar processo	Interpretação de desenho técnico; Metrologia; Processos de fabricação; Tecnologia de ferramentas de corte
E19	Executou acabamento superficial do rebaixo D com Ra 3,2	eixo com o rolamento montado	C8: Aplicar procedimentos e normas técnicas	Interpretação de desenho técnico; Metrologia; Processos de fabricação; Tecnologia de ferramentas de corte
E20	Utilizou os equipamentos de proteção individual (óculos de proteção, calçado de segurança e proteção para cabelos)	eixo com o rolamento montado	C8: Aplicar procedimentos e normas técnicas	Normas de saúde e segurança no trabalho e ambientais
E21	Utilizou os EPIs para realizar a manutenção (óculos de proteção, calçado de segurança e proteção para cabelos)	eixo com o rolamento montado	C8: Aplicar procedimentos e normas técnicas	Normas de saúde e segurança no trabalho e ambientais
E22	Identificou na tabela de Sistemas de Tolerâncias e Ajustes os desvios permissíveis	eixo com o rolamento montado	C1: Interpretar dados e informações	Ajustes e tolerâncias; Metrologia

CODIGO	EVIDÊNCIA	ATIVIDADE (ENTREGA)	CAPACIDADE	OBJETOS DE CONHECIMENTO
E23	Identificou na tabela técnica a classe de tolerância para montagem de rolamentos	eixo com o rolamento montado	C1: Interpretar dados e informações	Elementos de máquinas
E24	Selecionou dispositivo para a montagem	eixo com o rolamento montado	C3: Especificar recursos	Técnicas de manutenção (lubrificação e manutenção aplicada)
E25	Selecionou instrumento para medir a dimensão do alojamento	eixo com o rolamento montado	C3: Especificar recursos	Metrologia
E26	Selecionou EPI adequado conforme a atividade	eixo com o rolamento montado	C8: Aplicar procedimentos e normas técnicas	Normas de saúde e segurança no trabalho e ambientais
E27	Manteve o ambiente de trabalho organizado durante a realização da tarefa	eixo com o rolamento montado	C8: Aplicar procedimentos e normas técnicas	Normas de saúde e segurança no trabalho e ambientais
E28	Manteve a segurança e a integridade pessoal	eixo com o rolamento montado	C8: Aplicar procedimentos e normas técnicas	Normas de saúde e segurança no trabalho e ambientais
E29	Registrou a dimensão real do eixo na Ordem de Serviço conforme anexo 5	ordem de serviço preenchida e encerrada	C5: Controlar processo	Metrologia

CODIGO	EVIDÊNCIA	ATIVIDADE (ENTREGA)	CAPACIDADE	OBJETOS DE CONHECIMENTO
E30	Registrou dimensões do rolamento fornecido na Ordem de Serviço conforme anexo 5	ordem de serviço preenchida e encerrada	C5: Controlar processo	Metrologia
E31	Registrou as dimensões limites do eixo conforme tabela do sistema de tolerâncias e ajuste na Ordem de Serviço conforme anexo 5	ordem de serviço preenchida e encerrada	C1: Interpretar dados e informações	Ajustes e tolerâncias
E32	Encostou o anel interno do rolamento na face de encosto do eixo durante a montagem	eixo com o rolamento montado	C4: Definir técnicas	Técnicas de manutenção (lubrificação e manutenção aplicada)

ANEXO E - Termo de autorização para realização da pesquisa**SOLICITAÇÃO DE AUTORIZAÇÃO PARA PESQUISA ACADÊMICA**

Solicitamos autorização institucional para realização da pesquisa intitulada “**A avaliação do domínio psicomotor da aprendizagem: estudo para construção de escala de proficiência na educação profissional**” pelo acadêmico **Thiago Korb**, sob orientação do Professor Orientador **Dr. Dalton Francisco de Andrade**, para utilização da base de dados dos resultados da prova prática do SAEP 2017 – Técnico em Mecânica, idealizada pelo **Serviço Nacional de Aprendizagem Industrial (SENAI)**, com o objetivo de propor uma escala que permita avaliar na prática o domínio psicomotor de estudantes da educação profissional à luz da Teoria de Resposta ao Item.

Ao mesmo tempo, pedimos autorização para que o nome desta instituição conste no documento final, bem como futuras publicações e periódicos científico. Salientamos ainda que tais dados serão utilizados somente para a realização deste estudo ou serão mantidos em um banco de dados de pesquisa, com acesso restrito, para utilização em pesquisas futuras.

A presente atividade é requisito para a conclusão do Programa de Pós-graduação em Métodos e Gestão em Avaliação [Mestrado Profissional] da Universidade Federal de Santa Catarina (UFSC).

Na certeza de contarmos com a colaboração e empenho desta Gerência, agradecemos antecipadamente a atenção, ficando à disposição para quaisquer esclarecimentos adicionais que se fizerem necessários.

Brasília-DF, 14 de maio de 2019.

Thiago Korb
Acadêmico/Pesquisador
Universidade Federal de Santa Catarina

Luiz Eduardo Leão
Gerente de Tecnologias Educacionais
SENAI Departamento Nacional

Luiz Eduardo Leão
Gerente de Tecnologias
Educacionais

Hugo Nakatani
Analista de Desenvolvimento
Industrial

ANEXO F – Base de dados

A base de dados estava estruturada contemplando: máscara e ID dos estudantes, características observáveis (UF, sexo e idade) e o comportamento das respostas dos estudantes, sendo marcado 0 quando atribuído o “não” atendimento da evidência e 1 quando atribuído “sim”.

Máscara	ID_ Estudante	UF	Sexo	Idade	E01	E02	E03	E04	E05	E...	EV30	EV31	EV32	Acertos
10001	76337	AM	M	28	1	0	1	1	1	...	1	1	1	26
10002	76362	AM	M	20	1	1	1	1	1	...	1	1	1	32
10003	76372	AM	M	22	0	0	0	0	0	...	1	0	1	21
10004	76395	AM	M	23	1	1	1	1	1	...	1	1	1	29
10005	76425	AM	M	20	0	0	1	0	1	...	1	1	1	27
10006	76429	AM	F	28	0	0	1	0	1	...	0	1	1	25
10007	76436	AM	M	32	1	1	1	1	1	...	1	1	1	28
10008	77245	AM	M	26	1	1	1	1	1	...	1	1	1	31
10009	77284	AM	M	23	1	1	1	1	1	...	1	1	1	32
10010	77326	AM	M	33	1	1	1	1	1	...	1	1	1	31
10011	77708	AM	M	28	1	1	1	1	1	...	1	1	1	32
10012	87721	AM	M	31	1	1	1	1	1	...	1	1	1	31
10013	76687	AM	M	31	1	1	1	1	1	...	1	0	1	29

