

UNIVERSIDADE JOSÉ DO ROSÁRIO VELLANO - UNIFENAS

Pedro Paulo Trindade Resende

**AVALIAÇÃO DO TESTE DO PROGRESSO DE UMA FACULDADE DE MEDICINA
PELOS PRESSUPOSTOS DA TAXONOMIA SOLO**

**Belo Horizonte
2022**

Pedro Paulo Trindade Resende

**AVALIAÇÃO DO TESTE DO PROGRESSO DE UMA FACULDADE DE MEDICINA
PELOS PRESSUPOSTOS DA TAXONOMIA SOLO**

Dissertação apresentada ao curso de Mestrado Profissional em Ensino da Universidade José do Rosário Vellano - UNIFENAS para obtenção do título de Mestre em Ensino em Saúde.

Orientador: Prof. Dr. Alexandre Araújo Pereira

Coorientador: Prof. Dr. José Maria Peixoto

**Belo Horizonte
2022**

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Unifenas BH Itapoã

R433a Resende, Pedro Paulo Trindade

Avaliação do teste do progresso de uma faculdade de medicina pelo pressuposto da taxonomia solo. [manuscrito] / Pedro Paulo Trindade Resende. – Belo Horizonte, 2022.
57 f.

Orientador: Alexandre de Araújo Pereira.

Coorientador: José Maria Peixoto

Dissertação (Mestrado) – Universidade Professor Edson Antônio Velano, Programa de Pós-Graduação Stricto Sensu em Ensino em saúde, 2022.

1. Educação médica. 2. Aprendizagem. 3. Avaliação. I. Resende, Pedro Paulo Trindade. II. Universidade Professor Edson Antônio Velano. III. Título.

CDU: 61:378

Bibliotecária responsável: Gisele da Silva Rodrigues CRB6–2404

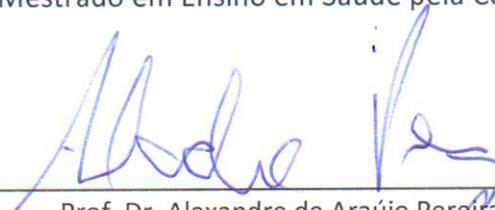
Certificado de Aprovação

AVALIAÇÃO DO TESTE DO PROGRESSO DE UMA FACULDADE DE MEDICINA PELOS PRESSUPOSTOS DA TAXONOMIA SOLO

AUTORA: Pedro Paulo Trindade Resende

ORIENTADORA: Prof. Dr. Alexandre de Araújo Pereira

Aprovado como parte das exigências para obtenção do Título de Mestre, no Programa de Pós-graduação Profissional de Mestrado em Ensino em Saúde pela Comissão Examinadora.



Prof. Dr. Alexandre de Araújo Pereira



Profa. Dra. Alice Belleigoli Rezende



Profa. Dra. Lígia Maria Cayres Ribeiro

Belo Horizonte, 19 de abril de 2023.



Prof. Dr. Antonio Carlos de Castro Toledo Jr.
Coordenador do Mestrado Profissional
Em Ensino em Saúde
UNIFENAS

RESUMO

Introdução: o preparo de futuros egressos das faculdades de medicina para uma prática assistencial responsável e qualificada de atenção à saúde é um grande desafio. O objetivo é que estejam aptos para a resolução de problemas complexos, que demandem habilidades cognitivas de ordem superior. Dessa forma, a avaliação da aquisição de tais competências se torna fundamental. Uma das formas de avaliação que vem sendo muito utilizada nos cursos de medicina é o Teste do Progresso (TP), que tem como uma de suas finalidades a verificação do desempenho cognitivo longitudinal dos estudantes durante o curso. Teorias cognitivas têm trazido avanços em pesquisas educacionais relativas à adequação dos processos de avaliação ao currículo, além de facilitar e orientar o princípio da progressividade do processo de ensino-aprendizagem. Usamos no nosso estudo a Taxonomia Structure of Observing Learning Outcome (SOLO) para avaliar e categorizar os itens do TP aplicado no primeiro semestre do ano de 2022 no curso de medicina da Universidade José do Rosário Vellano - Câmpus de Belo Horizonte (TPU2022-1). A Taxonomia SOLO (TS) permite a análise cognitiva necessária para a realização de determinadas tarefas, possibilitando uma observação integral da compreensão do entendimento do aluno, sendo composta por cinco níveis progressivos de complexidade de aprendizagem: pré-estrutural, uniestrutural, multiestrutural, relacional e abstrato estendido. Esses níveis de complexidade cognitiva são subdivididos em duas categorias, superficial e profunda. A aprendizagem superficial (AS) se baseia na retenção de detalhes factuais, através de memorização, e a aprendizagem profunda (AP) é relacionada com a organização e estruturação do conhecimento. Utilizamos também no nosso estudo a Teoria Clássica dos Testes (TCT) e calculamos para cada índice de múltipla escolha (IME) do TPU2022-1 o índice de dificuldade (IDF) e o índice de discriminação (ID) e os correlacionamos com a classificação SOLO. **Objetivo:** avaliar a qualidade do TPU2022-1, analisando seus itens pelos pressupostos da TS e os correlacionando com a TCT. **Materiais e Métodos:** estudo descritivo, exploratório, de abordagem quantitativa e qualitativa. De acordo com os princípios da TS, foi realizado a análise e caracterização dos itens da prova TPU2022-1 e a correlação com os IDF e ID. **Conclusão:** verificamos um equilíbrio entre a AS e a AP no total de itens e uma relação direta entre os níveis de AP e IME compostos por casos clínicos que exigiram habilidades de resolução de problemas, cognição e integração de conhecimento, fomentando um raciocínio clínico. Não verificamos diferença estatisticamente significativa entre as categorias SOLO quanto às médias do IDF e o ID.

Palavras-chave: Taxonomia; Educação em saúde; Aprendizagem; Avaliação; Teste do Progresso.

ABSTRACT

Introduction: The preparation of future graduates of medical schools for a responsible and qualified healthcare practice is a great challenge. The goal is to qualify them to solve complex problems that demand higher order cognitive skills. Therefore, the evaluation of the acquisition of such competencies becomes fundamental. One of the methods of evaluation that has been widely used in medical schools is the Progress Test (PT), which has as one of its purposes the verification of the longitudinal cognitive performance of students during the course. Cognitive theories have brought advances in educational research concerning the adequacy of evaluation methods to the curriculum, as well as facilitating and guiding the principle of progressivity in the teaching-learning process. In this study, the Structure of Observing Learning Outcome (SOLO) Taxonomy was used to evaluate and categorize the PT items applied in the first semester of 2022 in the medical course at the University José do Rosário Vellano - Belo Horizonte Campus (TPU2022-1). The SOLO Taxonomy (ST) allows for the cognitive analysis that is required to perform certain tasks, enabling a comprehensive observation of the student's understanding, and is composed of five progressive levels of learning complexity: prestructural, unistructural, multistructural, relational and extended abstract. These cognitive complexity levels are subdivided into two categories, superficial and deep. Superficial learning (SL) is based on the retention of factual details, through memorization, and deep learning (DL) is related to the organization and structuring of knowledge. The Classical Test Theory (CTT) was also used in our study and it was calculated for each multiple-choice index (MCI) of TPU2022-1 the difficulty index (DFI) and discrimination index (DI) and correlated them with the SOLO rating. **Objective:** to evaluate the quality of the TPU2022-1, analyzing its items by the assumptions of ST and correlating them with the CTT. **Materials and Methods:** descriptive, exploratory study with a quantitative and qualitative approach. In accordance with the principles of ST, the analysis and characterization of the items of the TPU2022-1 test and the correlation with the DFI and DI were performed. **Conclusion:** a balance was verified between SL and DL in the total items and a direct relation between DL and MCI levels composed of clinical cases that required problem-solving skills, cognition, and knowledge integration, fostering clinical reasoning. There was no statistically significant difference between the SOLO categories concerning the MCI and the DI mean scores.

Keywords: Taxonomy; Health education; Learning; Evaluation; Progress Test.

LISTA DE TABELAS

Tabela 1: Distribuição dos itens por área de conhecimento de TPU2022-1.	24
Tabela 2: Distribuição e classificação do grau de dificuldade dos itens em uma avaliação pela Teoria Clássica dos Testes.	26
Tabela 3: Classificação do índice de discriminação pela Teoria Clássica dos Testes	27
Tabela 4: Distribuição dos itens do TPU2022-1 entre aprendizagem superficial e aprendizagem profunda.	28
Tabela 5: Distribuição dos itens do TPU2022-1 pelos níveis da TS.	29
Tabela 6: Estratégia de formulação dos itens do TPU2022-1, indicando a presença ou ausência de casos clínicos.	31
Tabela 7: Estratégia de formulação dos itens do TPU2022-1, indicando a presença ou ausência de casos clínicos e a classificação SOLO.	31
Tabela 8: Índice de Dificuldade TPU2022-1.	32
Tabela 9: Índice de Discriminação TPU2022-1.	32
Tabela 10: Análises descritiva e comparativa entre as categorias do SOLO quanto aos parâmetros de avaliação do TCT (Dificuldade e Discriminação).	33
Tabela 11: Análises descritiva e comparativa entre duas categorias do SOLO quanto aos parâmetros de avaliação do TCT (Dificuldade e Discriminação).	35

LISTA DE QUADROS

Quadro 1: Exemplo de respostas classificadas pela Taxonomia SOLO.	14
Quadro 2: Sistema de categorização dos itens do TPU2022-1.	25

LISTA DE FIGURAS

Figura 1: Verbos utilizados para formular objetivos em cada nível da TS.....	15
--	----

LISTA DE GRÁFICOS

Gráfico 1: BOXPLOT do parâmetro IDF, por SOLO.....	34
Gráfico 2: BOXPLOT do parâmetro ID, por SOLO	34
Gráfico 3: BOXPLOT do parâmetro “Dificuldade”, por SOLO (2 categorias)	35
Gráfico 4: BOXPLOT do parâmetro “Discriminação”, por SOLO (2 categorias).....	36

LISTA DE ABREVIATURAS, SÍMBOLOS E SIGLAS

ABEM	Associação Brasileira de Educação Médica
AP	Aprendizagem profunda
APB	Aprendizagem Baseada em Problemas
AS	Aprendizagem superficial
CEP	Comitê de Ética em Pesquisa
DCN	Diretrizes Curriculares Nacionais
DM	Diabetes mellitus
FCI	Ficha de categorização do item
HAS	Hipertensão arterial sistêmica
ID	Índice de discriminação
IDF	Índice de dificuldade
IME	Índice de múltipla escolha
mini-CEX	Miniexercício de avaliação clínica
OSCE	<i>Objective Structured Clinical Examination</i> [Exame Clínico Objetivo Estruturado]
PAAF	Punção aspirativa por agulha fina
SOLO	<i>Structure of Observing Learning Outcome</i>
SUPREMA	Faculdade de Ciências Médicas e da Saúde de Juiz de Fora
TA	Termo de Anuência
TB	Taxonomia de Bloom
TCLE	Termo de Consentimento Livre e Esclarecido
TCT	Teoria Clássica dos Testes
TCUD	Termo de Compromisso para Utilização dos Dados
TEP MINAS	Consórcio Mineiro de Escolas Médicas para o Teste de Progresso
TM	Taxonomia de Marzano
TP	Teste do Progresso

TPU2022-1	Teste do Progresso UNIFENAS-BH 2022 1
TS	Taxonomia SOLO
UFJF	Universidade Federal de Juiz de Fora
UFOP	Universidade Federal de Ouro Preto
UFSJ	Universidade Federal de São João del-Rei
UFTM	Universidade Federal do Triângulo Mineiro
UFV	Universidade Federal de Viçosa
UNIFENAS	Universidade José do Rosário Vellano

SUMÁRIO

1	INTRODUÇÃO	10
1.1	<i>Taxonomias cognitivas.....</i>	<i>11</i>
1.2	<i>Avaliação de aprendizagem.....</i>	<i>17</i>
1.3	<i>Teste do Progresso.....</i>	<i>18</i>
2	JUSTIFICATIVA	21
3	OBJETIVOS	22
3.1	<i>Objetivo geral.....</i>	<i>22</i>
3.2	<i>Objetivos específicos.....</i>	<i>22</i>
4	MATERIAIS E MÉTODOS	23
4.1	<i>Desenho do estudo.....</i>	<i>23</i>
4.2	<i>Contexto de realização do estudo.....</i>	<i>23</i>
4.3	<i>Critérios de escolha das provas.....</i>	<i>24</i>
4.4	<i>Coleta de dados.....</i>	<i>24</i>
4.5	<i>Plano de análise das questões.....</i>	<i>24</i>
4.6	<i>Análise Psicométrica pela Teoria Clássica dos Testes.....</i>	<i>26</i>
4.7	<i>Análise estatística.....</i>	<i>27</i>
4.8	<i>Aspectos éticos.....</i>	<i>27</i>
5	RESULTADOS.....	28
5.1	<i>Análise da complexidade cognitiva dos itens.....</i>	<i>28</i>
5.2	<i>Análise Psicométrica pela Teoria Clássica dos Testes.....</i>	<i>32</i>
5.2.1	<i>Índice de Dificuldade.....</i>	<i>32</i>
5.2.2	<i>Índice de Discriminação.....</i>	<i>32</i>
5.3	<i>Correlação entre os Índices de Dificuldade e de Discriminação com a Classificação SOLO.....</i>	<i>33</i>
6	DISCUSSÃO	37
7	CONCLUSÃO	45
8	REFERÊNCIAS	46
	APÊNDICE A - Solicitação de dispensa do Termo de Consentimento Livre e Esclarecido (TCLE)	51
	APÊNDICE B - Termo de Compromisso de Utilização de Dados (TCUD)	52

APÊNDICE C - Ficha de categorização do item (FCI).....	54
APÊNDICE D - Sistema de categorização dos itens pela Taxonomia SOLO	55
APÊNDICE E - Tabelas SOLO	56
ANEXO A - Link de acesso aos itens do TPUBH2022-1.....	57

1 INTRODUÇÃO

O grande objetivo das escolas médicas é preparar futuros egressos para uma prática assistencial responsável e qualificada de atenção à saúde, que atenda às necessidades de uma sociedade contemporânea. As Diretrizes Curriculares Nacionais (DCN) do Curso de Graduação em Medicina definem o egresso da faculdade de medicina com formação generalista, humanista, crítica, reflexiva e ética (BRASIL, 2014).

Nesse sentido, o processo de avaliação da aprendizagem é fundamental, pois permite observar os efeitos das intervenções educacionais, além de informar se o ciclo de aprendizagem percorrido está proporcionando a construção de uma aprendizagem profunda, aquela capaz de mobilizar conhecimentos, habilidades e atitudes para a resolução dos problemas complexos (CHAMPLAIN, 2010). A avaliação da aprendizagem é também uma ferramenta impulsionadora no processo de ensino, porém muitas vezes pouco explorada para avaliar habilidades de pensamento de ordem superior, domínios afetivos e psicomotores, ficando restrita a aspectos cognitivos de baixa ordem (HADIE, 2018).

Um método de avaliação muito utilizado nas escolas de medicina nos dias de hoje é o Teste do Progresso (TP), que analisa o desempenho cognitivo longitudinal dos estudantes, baseando-se nas competências cognitivas esperadas no final do curso (REBERTI *et al.*, 2020). O TP aplicado no primeiro semestre do ano de 2022 no Curso de Medicina da Universidade José do Rosário Vellano - Câmpus de Belo Horizonte (TPU2022-1) é uma avaliação aplicada a todos os alunos do curso, do 1º ao 12º período, contendo 120 itens de múltipla escolha (IME), sendo o conteúdo dividido pelas grandes áreas do conhecimento: cirurgia, clínica médica, ginecologia-obstetrícia, pediatria e saúde coletiva.

No propósito de saber o que realmente está sendo avaliado no TPU2022-1, analisamos a prova através de uma ferramenta educacional taxonômica, que orienta a elaboração dos itens de uma avaliação. Taxonomia é um sistema de classificação que, no campo educacional, categoriza os níveis de aprendizagem atingidos pelos estudantes, sendo útil para o sistema de avaliação e para o planejamento dos objetivos educacionais (ARIPIN *et al.*, 2020). Teorias cognitivas com foco neopiagetiano têm fornecido compreensão sobre a evolução do conhecimento humano em estágios do desenvolvimento cognitivo (FERREIRA; ROCHA, 2020).

Esse estudo avaliou a complexidade cognitiva dos itens de um TP realizado em um curso de medicina brasileiro pelos pressupostos de uma taxonomia denominada *Structure of Observing Learning Outcome* (SOLO). Através da análise da complexidade cognitiva necessária para resolução de seus itens, podemos avaliar qualitativamente a aprendizagem (SPRECHER, 2019).

Avaliamos também, através da Teoria Clássica dos Testes (TCT), o TPU 2022-1, e calculamos para cada item o índice de dificuldade (IDF) assim como o índice de discriminação (ID) e os correlacionamos com a classificação SOLO. A TCT compreende métodos que dão subsídios para a concepção de instrumentos de medidas em pesquisas, buscando justificar o sentido das respostas dos alunos nos processos de avaliação (SARTES; SOUZA-FORMIGONI, 2013).

1.1 Taxonomias cognitivas

As taxonomias cognitivas trouxeram grande auxílio no processo de ensino e aprendizagem, podendo ser utilizadas como modelos hierárquicos para classificação de metas ou objetivos educacionais de acordo com níveis de complexidade de entendimento (ARIPIN *et al.*, 2020).

Uma das mais conhecidas é a Taxonomia de Bloom (TB), tendo a característica de classificar objetivos educacionais e auxiliar o alcance desses em diferentes contextos (FERRAZ; BELHOT, 2010). A TB, apresentada em 1956 e revisada em 2001, classifica hierarquicamente os níveis de aprendizagem, do pensamento mais básico ao mais elaborado (BORNEA; GONÇALVES; PADOVANI, 2014). Os objetivos de aprendizagem são mais fáceis de serem atingidos a partir do desenvolvimento cognitivo linear, de conceitos mais básicos para os mais elaborados. Através da utilização de verbos de ação, que descrevem os processos cognitivos desejados, podemos especificar os objetivos a serem alcançados (FERRAZ; BELHOT, 2010).

A Taxonomia de Marzano (TM) é baseada em duas dimensões: processamento de informação e nível de conhecimento necessário. Essa taxonomia incorpora outros fatores relacionados à aprendizagem, sendo composta por três sistemas: interno, metacognitivo e cognitivo. Perante uma nova tarefa, o sistema interno é o primeiro a se envolver com a decisão para a sua execução. Em seguida, o sistema metacognitivo discerne metas e estratégias para que o sistema cognitivo execute ações necessárias para sua realização. Enquanto na TB o domínio cognitivo é o mais utilizado, a TM adiciona a utilização da metacognição e o autossistema (IRVINE, 2017).

A Taxonomia Depth-of-Knowledge avalia categorias de profundidade de conhecimento, classificando em quatro níveis de complexidade ascendente. O nível 1 requer dos alunos

reprodução da memorização. O nível 2 exige algum processamento mental, mas ainda de forma superficial. O nível 3 envolve um raciocínio abstrato e complexo, porém restrito à atividade em questão. O nível 4, mais do que exigir um conhecimento profundo, vai além da atividade requerida, como formação de análises e sínteses (WEAY; MASOOD; ABDULLAH, 2016).

A Taxonomia SOLO (TS) é uma ferramenta que também avalia a qualidade da aprendizagem dos estudantes (MOL; MATOS, 2019). Trata-se de um instrumento, influenciado pelas teorias de Jean Piaget, biólogo e psicólogo, que classificou o desenvolvimento cognitivo em quatro estágios: sensório-motor, pré-operatório, operatório concreto e operatório formal. O período sensório-motor vai do nascimento aos dois anos de idade, quando a criança se relaciona com o mundo através de sensações e movimentos. No período pré-operatório, que geralmente vai dos dois aos sete anos de idade, a criança usa símbolos, números e palavras para relacionar-se, tendo uma atitude egocêntrica. O período operatório concreto é caracterizado pelo início da realização de operações mentais lógicas, normalmente indo dos sete aos onze anos. O último período do desenvolvimento é o operatório formal, que se inicia na adolescência, permanecendo até a idade adulta. Esse estágio é caracterizado pelo pensamento crítico, planejamento e raciocínio dedutivo (SCOTT; COGBURN, 2021).

Uma das limitações para a aplicação da Teoria de Piaget em adultos é o fato de que nessa fase da vida já houve o desenvolvimento de todos os estágios cognitivos. Procurando aprimorar as ideias de Piaget, teóricos neopiagetianos defendem que as estruturas cognitivas continuam ao longo da vida adulta, sendo ativamente criadas e recriadas pelos estudantes, tornando-se progressivamente mais complexas pela interação da maturação à experiência. Esses teóricos propõem que a complexidade da estrutura cognitiva evolui em estágios qualitativamente diferentes, nos quais os níveis mais abstratos e complexos são construídos a partir dos níveis inferiores (KNIGHT; SUTTON, 2004).

A partir dessa teoria surge a TS, que possui duas características principais: modo de pensar, relacionados à teoria de Piaget (sensório-motor, pré-operatório, operatório concreto e operatório formal), e o nível de resposta, que é o nível crescente de complexidade das respostas dos alunos. A TS objetiva classificar a complexidade da estrutura da aprendizagem demonstrada pelos estudantes, partindo da análise das respostas fornecidas por estes a tarefas específicas, não se preocupando com o modo de funcionamento cognitivo geral (MOL, 2019). Permite avaliar as respostas dos alunos em profundidade, qualidade e estrutura, sendo uma importante ferramenta no processo de avaliação (KORKMAZ; UNSAL, 2017).

A TS tem característica hierárquica, ou seja, cada nível mantém as competências adquiridas nos níveis anteriores. Para que um indivíduo exerça uma tarefa mais complexa, é necessário que seja capaz de realizar tarefas de menor complexidade no mesmo contexto (BIGG; COLLIS, 1982). Possui também a propriedade de flexibilidade, permitindo uma análise de determinada situação (YURTYAPAN; YILMAZ, 2021). Ordena os resultados de aprendizagem de acordo com a complexidade, possibilitando uma análise integral da compreensão do entendimento dos alunos e pode ser utilizada em qualquer disciplina ou grau de escolaridade (JIMOYIANNIS, 2011).

Para que haja uma potencialização do aprendizado, deve existir um alinhamento em todas as partes do ambiente de aprendizagem, como currículo, avaliação e resultado. A TS permite esse alinhamento (ILGUY *et al.*, 2014). No nível mais alto da TS, o aluno se torna capaz de emitir informações relevantes além daquelas ensinadas pelo educador. Na resposta menos madura, verifica-se pouco esclarecimento e inconsistência (SPRECHER, 2019).

A TS classifica as respostas fornecidas pelos estudantes em cinco níveis progressivos de complexidade cognitiva:

- Pré-estrutural: as respostas são inadequadas, apresentando informações irrelevantes ou incoerentes.
- Uniestrutural: o foco é adequado, mas o aluno se detém a um único elemento relevante e as respostas ficam inconsistentes.
- Multiestrutural: o aluno se vale de mais informações, mas elas não se integram e incoerências podem aparecer nas suas respostas.
- Relacional: as informações são incorporadas e estabelecidas relações de forma coerente, havendo um entendimento do todo, não havendo inconsistências.
- Abstrato estendido: a resposta do aluno vai além dos elementos da questão, partindo para abstração e generalização.

O quadro 1 apresenta cinco exemplos de respostas a um problema clínico de um paciente com complicações secundárias ao diabetes mellitus (DM) e a hipertensão arterial sistêmica (HAS), de acordo com as categorias da TS.

Quadro 1: Exemplo de respostas classificadas pela Taxonomia SOLO.

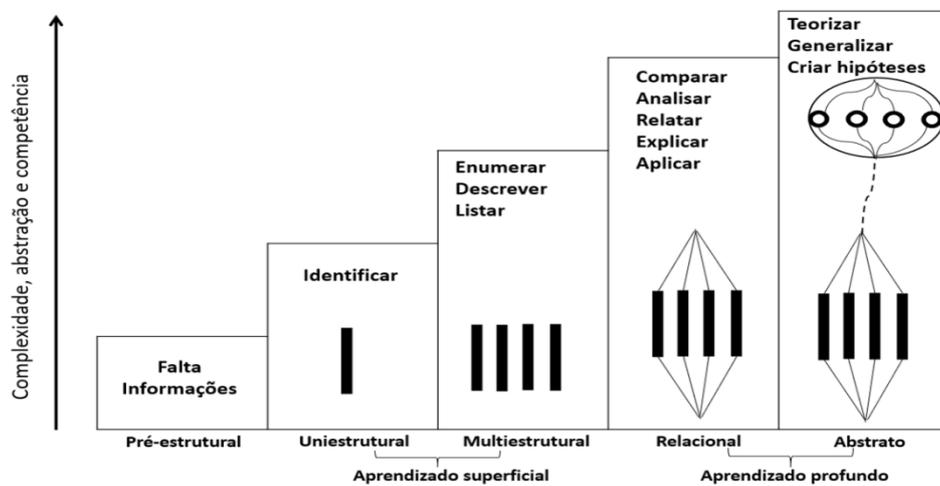
Problema: Paciente com 43 anos de idade, portador de DM e HAS há mais de 20 anos, apresentando disfunção renal e perda da acuidade visual. Informa uso irregular da medicação e das orientações dietéticas, com controle inadequado das duas patologias. Qual a explicação fisiopatológica para os problemas observados no paciente?	
Pré-estrutural	A idade do paciente
Uniestrutural	O DM
Multiestrutural	O DM e a HAS
Relacional	O DM é uma doença que, se não devidamente controlada, pode levar a complicações microvasculares crônicas, como a nefropatia e a retinopatia através de mecanismos metabólicos e hemodinâmicos. Essas complicações podem ser potencializadas pela HAS com controle insatisfatório.
Abstrato estendido	Além da explicação fisiopatológica de lesões microvasculares e apesar de não relatados sintomas cardiovasculares, o paciente diabético e hipertenso com inadequada adesão terapêutica e longo prazo de instalação dessas doenças apresenta risco cardiovascular aumentado pela maior prevalência de acometimento macrovascular, elevando a possibilidade de ocorrência de infarto do miocárdio, acidente vascular encefálico e morte súbita.

O nível pré-estrutural apresenta uma resposta incorreta, inadequada, que não explica o problema do paciente, confundido por um aspecto irrelevante. A idade do paciente não justifica a fisiopatologia do caso. A resposta uniestrutural traz um fato relevante ao afirmar que o DM poderia explicar os problemas do paciente, mas não analisa outros dados e, portanto, trata-se de uma resposta inconsistente. Na resposta multiestrutural, o aluno identifica que o DM e a HAS poderiam explicar os problemas do paciente, mas não estabelece nenhuma relação, como potencialização dessas doenças na incidência de possíveis complicações. Já no nível relacional, o estudante relaciona uma provável associação dessas duas patologias na prevalência e potencialização de complicações, como nefropatia e retinopatia, através de mecanismos metabólicos e hemodinâmicos, além de citar a não adesão terapêutica e o longo tempo de evolução das doenças. No nível abstrato estendido, existe, além da relação dos tópicos na fisiopatologia da doença, inferência a respeito de complicações ainda não presentes, como o infarto do miocárdio, o acidente vascular encefálico e a morte súbita.

Uma forma que auxilia a classificação de itens pelos níveis de complexidade da TS é a análise do verbo de comando utilizado na formulação das perguntas. No nível uniestrutural, a pergunta requer um único aspecto relevante para a sua realização. Nesse nível, as ações utilizadas são: identificar ou realizar um procedimento simples. No nível multiestrutural, mais de uma

informação é necessária para sua resposta, porém não existe integração entre elas. “Enumerar”, “descrever”, “listar” e “classificar” são os verbos relacionados a esse nível. O nível relacional também exige mais de um elemento para sua realização, no entanto existe uma integração coerente entre estes. Os verbos relacionados com esse nível são: “comparar”, “contrastar”, “explicar causas”, “analisar”, “relacionar” e “aplicar”. No nível abstrato estendido, as questões exigem, além de mais elementos para sua resposta, um alto nível de integração e generalização. Os verbos relacionados a esse nível são: “teorizar”, “generalizar”, “formalizar hipóteses” e “refletir” (FERREIRA; ROCHA, 2020). A figura 1 relaciona os verbos ao nível SOLO.

Figura 1: Verbos utilizados para formular objetivos em cada nível da TS.



Fonte: Traduzido de Biggs e Tang (2007).

Apesar do potencial existente na análise do verbo de comando utilizado no enunciado de um item para a classificação pela TS, deve-se atentar para o fato de que alguns verbos podem pertencer a mais de um nível taxonômico, dependendo do contexto empregado (SCULLY, 2017).

Os níveis de complexidade cognitiva descritos por Biggs e Collis (1982) são subdivididos em duas categorias: aprendizagem superficial (AS) e aprendizagem profunda (AP). A AS é formada pelos níveis uniestrutural e multiestrutural; a AP, pelos níveis relacional e abstrato estendido. A AS se baseia na retenção de detalhes factuais, através de memorização em curto prazo, não exigindo relações de conhecimentos prévios e atuais, ao passo que a AP é mais complexa, requerendo relações entre os dados, podendo capacitar o indivíduo para entender mecanismos e princípios e realizar generalizações ou teorizações (ROSSI *et al.*, 2021).

Ganha grande importância um arranjo entre os itens de uma avaliação de acordo com as taxonomias de aprendizagem. Não se deseja o uso exclusivo de itens relacionados à AP, mas que os alunos sejam experimentados em todos os níveis da TS (LISTER *et al.*, 2006). Para que um aluno alcance a AP, é necessário que ele obtenha primeiramente a AS (BIGG; COLLIS, 1982). Hattie e Brown (2004) recomendam um equilíbrio entre os níveis de complexidade superficial e profunda nos itens que compõem uma avaliação, contemplando, pelo menos, 30% das questões a cada um desses níveis.

Além de ser um instrumento de identificação da complexidade de desempenho dos estudantes na realização de determinadas tarefas, a TS tem sido também utilizada com outros objetivos, como classificação de questões de acordo com a complexidade cognitiva, avaliação da qualidade de aprendizagem e de programas de ensino (MOL; MATOS, 2019).

Ao analisar as questões pela TS, verifica-se que o grau de complexidade não é determinante do seu nível de dificuldade (HATTIE; BROWN, 2004; SCULLY, 2017). Dessa forma, é necessário diferenciar o grau de complexidade cognitiva exigida por uma questão – que se refere às operações de pensamento necessárias para sua resolução – do grau de dificuldade, que se relaciona a fatores que oferecem obstáculos para a resolução, como, por exemplo, a familiaridade com o assunto. Podem existir questões com maior nível de dificuldade e menor nível de complexidade e vice-versa. O grau de dificuldade das questões é avaliado por métodos estatísticos, ao passo que o nível de complexidade pode ser determinado por metodologias qualitativas, como a TS (PEREIRA, 2019).

Um item de maior complexidade, como a resolução de um caso clínico, pode ser facilmente respondido se o aluno teve contato com casos semelhantes nas suas atividades práticas. Porém, o mesmo aluno pode ter um maior obstáculo na resolução de uma questão de menor dificuldade, como responder qual agente implicado na fisiopatologia de uma doença infecciosa, se porventura ele não teve contato, teórico ou prático, com a respectiva enfermidade.

A TS foi escolhida para a avaliação no nosso estudo, pois permite a análise e a categorização dos níveis de aprendizagem demonstrados pelos alunos em uma tarefa específica, avaliando e identificando os processos de pensamento utilizados para a sua resolução (YURTYAPAN; YILMAZ, 2021). Permite também uma diferenciação mais concludente entre a aprendizagem superficial e a profunda. Scully (2017) cita uma certa inconsistência na TB para o discernimento de onde termina o pensamento de ordem inferior e de onde inicia o de ordem superior.

1.2 Avaliação de aprendizagem

O processo de avaliação se tornou uma das principais ferramentas no arsenal do ensino médico, merecendo cuidadosa atenção dos educadores e sendo um instrumento impulsionador de aprendizagem (O'SHAUGHNESSY; JOYCE, 2015; FERRIS; O'FLYNN, 2015; PRASHANTI; RAMNARAYAN, 2019). Para que haja uma melhor aprendizagem deve-se categorizar e objetivar os diferentes níveis de habilidades cognitivas, afetivas e domínios psicomotores desejáveis a serem alcançados pelos discentes (HADIE, 2018).

O critério de aprovação nos processos de avaliação das escolas médicas é de grande preocupação, pois reflete o domínio de competências necessárias para o exercício da profissão de maneira eficaz e segura ao paciente e a sociedade (CHAMPLAIN, 2010). Devem ser selecionados métodos de avaliação adequados, em validade e confiabilidade, para identificar os três objetivos a serem atingidos nas escolas de medicina: conhecimentos, habilidades e atitudes (FOWELL; BLIGH, 1998).

Com relação ao momento em que as avaliações são realizadas, bem como aos seus objetivos principais, elas podem ser definidas como diagnósticas, somativas ou formativas. A avaliação diagnóstica, que é realizada no princípio de um módulo ou período, tem a finalidade de verificar os conhecimentos prévios dos alunos. Isso permite uma melhor estruturação das atividades educacionais. A avaliação somativa é a “avaliação da aprendizagem”, que normalmente é realizada no final de um módulo ou disciplina, tendo como finalidade quantificar o nível de competências adquiridas pelos alunos. A avaliação formativa, chamada “avaliação para aprendizagem”, fornece um processo contínuo de *feedback* e identifica os pontos de aprendizagem alcançados pelos estudantes, detectando necessidades de aprimoramento, sendo uma valiosa ferramenta de gestão acadêmica, fornecendo importantes informações para a gestão do currículo escolar (ANDREASSEN; MALLING, 2019).

Todas as formas de avaliação, diagnóstica (avaliando conhecimento prévio), formativa (acompanhando e direcionando continuamente o processo de aprendizagem) e somativa (garantindo um padrão mínimo de aquisição de competências), são importantes e necessárias no currículo das escolas médicas quando utilizadas no momento adequado e de forma oportuna. (FERRIS; O'FLYNN, 2015; SCHUTTPELZ-BRAUNS *et al.*, 2020).

Neste estudo, analisamos itens de um instrumento de avaliação muito utilizado no ensino médico, que é a prova com índice de múltipla escolha (IME). Isso acontece devido às vantagens

oferecidas por esse método, como possibilidade de avaliação dos diversos níveis de complexidade do domínio cognitivo, aplicação a um grande número de alunos, abrangência de um grande conteúdo, facilidade de correção, com possibilidade de oferecimento de devolutiva de maneira mais rápida (BOLLELA; BORGES; TROCON, 2018). Quando realizada de forma on-line, tem a vantagem de ser personalizável, oferecendo opções quanto ao tempo de realização, apresentação e sequenciamento dos itens e disponibilizando resultados e *feedback* de maneira instantânea (DAVIES; PROOPS; CAROLAN, 2020).

Críticos argumentam sobre a dificuldade da avaliação do aprendizado profundo através de IME, estando este mais associado ao aprendizado superficial. Isso ocorre, muitas vezes, não devido a seu formato, mas sim à sua má construção, utilizando verbos que exigem aprendizagem de ordem inferior, como “memorizar” e “combinar” (DAVIES; PROOPS; CAROLAN, 2020; JAVAIED, 2021). Os IMEs, quando bem elaborados, possuem validade para analisar a aquisição de competências cognitivas superiores (SHAH *et al.*, 2012). Uma forma de alcançar esse objetivo seria forçando os alunos na resolução de problemas clínicos que demandam o pensamento crítico, que exigiria um raciocínio de alto nível. Raciocínio clínico é o processo cognitivo que interpreta informações, desenvolve hipóteses diagnósticas para colocar em prática condutas terapêuticas (ARAGÃO; ALMEIDA, 2017).

Os IMEs utilizados em uma avaliação devem ser analisados após a sua realização para a verificação de sua qualidade. A análise psicométrica pela TCT permite informações sobre a confiabilidade e a validade dos IMEs através do cálculo do IDF e do ID e sua inter-relação. Essa análise permite uma seleção de IMEs que avaliem as competências desejáveis e o aperfeiçoamento de itens não satisfatórios (GAJJAR *et al.*, 2014; BHATTACHERJEE *et al.*, 2022).

Outros métodos também são incorporados aos processos de avaliação do ensino médico, como miniexercício de avaliação clínica (mini-CEX), exame clínico objetivo estruturado (OSCE), simulações, grupos tutoriais, portfólios reflexivos, autoavaliação, avaliação por pares e itens dissertativos (FERRIS; O'FLYNN, 2015). Uma estratégia de avaliação que vem sendo utilizada pelas escolas médicas no mundo inteiro é o TP (SAKAI *et al.*, 2008).

1.3 Teste do Progresso

O TP é uma importante ferramenta de avaliação de ensino, tendo sido desenvolvido na Universidade de Missouri-Kansas City School of Medicine (EUA) e na Universidade de

Limburg (Holanda) no início dos anos 1970, sendo atualmente amplamente utilizado no mundo inteiro (GOLD *et al.*, 2015).

O objetivo principal do TP é avaliar o desempenho cognitivo longitudinal dos estudantes durante o curso. Tem também outras importantes utilidades, como propiciar ao estudante uma oportunidade de autoavaliação, avaliação do seu desempenho em relação a seus pares, treinamento para os concursos de residência médica, além de ser uma valiosa ferramenta de gestão acadêmica, fornecendo importantes informações para a gestão curricular (PINHEIRO *et al.*, 2015).

O TP é aplicado periodicamente (semestralmente ou anualmente) a todos os alunos do curso, de maneira simultânea, sendo constituído de IMEs e seu conteúdo compreende todo currículo e é relacionado aos seus objetivos finais e baseado nas Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina. Dessa forma, o TP deve exigir dos discentes muito além da memorização, ou seja, a utilização de um conhecimento mais aprofundado para a sua resolução. Seus itens são elaborados por professores da própria instituição ou de consórcios formados com esse objetivo e espera-se um ganho cognitivo progressivo com o passar dos períodos (BALDIM; VICENTE; ARCUR, 2018; ROSA *et al.*, 2017; PINHEIRO *et al.*, 2015).

O TP é uma das formas de avaliação externa do Curso de Medicina da Universidade José do Rosário Vellano - Câmpus de Belo Horizonte, que o realiza desde o ano de 2011, tendo este sido composto por 60 IMEs e ampliado para 120 itens nos anos de 2012 e 2013. Nos anos de 2014, 2016, 2017, 2018 e 2019 o curso de medicina da UNIFENAS optou por integrar o Consórcio Mineiro de Escolas Médicas para o Teste de Progresso (TEP MINAS), tendo o TP sido elaborado por esse consórcio e contendo 120 itens. Fazem parte do TEP MINAS as seguintes escolas médicas: Faculdade de Ciências Médicas e da Saúde, de Juiz de Fora (SUPREMA), Universidade Federal de Juiz de Fora (UFJF), Universidade Federal de Ouro Preto (UFOP), Universidade Federal de Viçosa (UFV), Universidade Federal de São João del-Rei (UFSJ), Universidade Federal do Triângulo Mineiro (UFTM) e Universidade José do Rosário Vellano (UNIFENAS) (Câmpus Alfenas e Belo Horizonte). Em 2015, a UNIFENAS participou de um TP nacional, envolvendo vários cursos médicos em todo o Brasil, coordenado pela Associação Brasileira de Educação Médica (ABEM). No ano de 2021 e no primeiro semestre de 2022 a UNIFENAS realizou o TP novamente de forma individualizada. No segundo semestre do ano de 2022, a UNIFENAS realizou outro TP, mais uma vez integrando o TEP MINAS.

O TPU2022-1 foi constituído de 120 IMEs, contendo quatro alternativas de resposta, com uma única opção correta. O conteúdo é dividido pelas grandes áreas do conhecimento: cirurgia, clínica médica, ginecologia-obstetrícia, pediatria e saúde coletiva. A nota final de cada aluno no TPU2022-1 substitui uma das formas de avaliação do curso de medicina da UNIFENAS-BH, o Exame Geral Integrado.

2 JUSTIFICATIVA

A finalidade do ensino médico é preparar os formandos para o exercício criterioso da profissão. A análise da aquisição do domínio de competências necessárias para o desempenho dessa atividade de maneira eficaz e segura ao paciente e à sociedade é fundamental.

Teorias cognitivas com foco neopiagetiano, como a TS, têm dado substrato para metodologias de pesquisas na avaliação da qualidade da aprendizagem, identificando níveis progressivos de complexidade do entendimento.

O TP é uma ferramenta de avaliação que tem por finalidade a análise do desempenho cognitivo dos estudantes, sendo também um instrumento de autoavaliação, além de fornecer importantes informações para a gestão curricular.

Apesar de sua relevância como instrumento de avaliação, poucos estudos analisando os níveis de complexidade cognitiva exigidos pelo TP de um curso de medicina têm sido publicados. Esse é o objetivo deste estudo, que analisou o TPU2022-1 e categorizou seus itens em relação à complexidade cognitiva pelos pressupostos da TS. Também objetivou correlacionar os IDF e ID com a classificação SOLO.

Espera-se que os resultados do presente estudo contribuam para o conhecimento das características dos itens que compõem o TPU2022-1, bem como possibilitem seu aprimoramento no sentido de cumprir seu objetivo no processo de avaliação.

3 OBJETIVOS

3.1 Objetivo geral

Avaliar a qualidade do TP aplicado no curso de medicina da UNIFENAS-BH no primeiro semestre do ano de 2022, analisando seus itens pelos pressupostos da TS e correlacionando-os com a TCT.

3.2 Objetivos específicos

Categorizar os itens do TP aplicado no curso de medicina da UNIFENAS-BH realizado no primeiro semestre do ano de 2022 pelos pressupostos da TS.

Avaliar a distribuição dos itens do TP aplicado no curso de medicina da UNIFENAS-BH no primeiro semestre do ano de 2022 entre as categorias da TS.

Comparar a classificação SOLO entre itens do TP aplicado no curso de medicina da UNIFENAS-BH no primeiro semestre do ano de 2022 que contém ou não casos clínicos.

4 MATERIAIS E MÉTODOS

4.1 *Desenho do estudo*

Trata-se de um estudo descritivo, exploratório, de abordagem quantitativa e qualitativa.

4.2 *Contexto de realização do estudo*

O curso de medicina da Universidade José do Rosário Vellano - Câmpus de Belo Horizonte teve início do seu funcionamento no primeiro semestre de 2003, em regime semestral e modular, do 1º ao 12º período. A UNIFENAS adotou no curso de medicina de Belo Horizonte a Aprendizagem Baseada em Problemas (ABP) como perspectiva educacional central. Outros pilares sobre os quais se construiu o currículo do curso de medicina da UNIFENAS Câmpus de Belo Horizonte são: o ensino centrado no estudante; baseado em aprendizagem contextualizada, em uma abordagem temática multidisciplinar e integrada; fundamentado e baseado na comunidade; atento às necessidades do SUS, em conformidade com as Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina (Resolução CNE/CES nº 3, de 20/06/2014); e priorizando a formação de médicos generalistas.

O TPU2022-1 foi constituído de 120 itens. Todos os itens foram de múltipla escolha, contendo quatro alternativas de resposta, com uma única opção correta. O conteúdo é dividido pelas grandes áreas do conhecimento: cirurgia, clínica médica, ginecologia-obstetrícia, pediatria e saúde coletiva. Trata-se de uma avaliação aplicada a todos os alunos do curso, do 1º ao 12º período, de forma simultânea, com tempo de duração de quatro horas para sua realização. Um *feedback* com gabarito comentado e a nota de cada aluno foram liberados após a prova.

Para o planejamento da elaboração do TPU2022-1, foi utilizada a matriz do TEP MINAS do ano de 2019. A matriz continha as 120 “encomendas” (itens) elaboradas pelos professores das escolas médicas associadas ao TEP MINAS que compreendia o tema da questão, o nível de atenção, a idade, o sexo do paciente e a habilidade requerida (planejamento do diagnóstico, anamnese ou exame físico, planejamento terapêutico, etc.). Após a realização, os itens foram avaliados por uma comissão do TEP MINAS. Também foram realizadas uma seleção de questões e a execução de aperfeiçoamentos, quando necessários.

A tabela 1 apresenta a distribuição das questões do TPU2022-1, de acordo com as áreas de conhecimento e o número de itens.

Tabela 1: Distribuição dos itens por área de conhecimento de TPU2022-1.

Ano/Área	Clínica Médica	Cirurgia	Ginecologia obstetrícia	Saúde Pública	Pediatria	Nº total
2022	24	24	24	24	24	120

4.3 Critérios de escolha das provas

Utilizamos neste estudo a prova do TPU2022-1, uma vez que o objetivo da pesquisa foi avaliar um exame atual e realizado exclusivamente na UNIFENAS-BH.

4.4 Coleta de dados

Solicitamos à coordenação do Núcleo de Avaliação do Curso de Medicina da UNIFENAS de Belo Horizonte a prova do TPU2022-1, incluindo texto-base, enunciado, alternativas, gabarito e comentário pedagógico de cada item. Solicitamos também a prova do TEP MINAS 2018 para a realização de um projeto piloto. Um link contendo as análises e as categorizações dos itens do TPU2022-1 está disponível no anexo A.

4.5 Plano de análise das questões

Todos os itens do TPU2022-1 foram analisados de acordo com os princípios da TS. As questões foram categorizadas em quatro níveis de complexidade cognitivas: uniestructural, multiestructural, relacional e abstrato estendido. O nível pré-estrutural da TS não foi incluído, uma vez que o propósito deste estudo foi analisar as perguntas das questões, objetivando verificar a complexidade cognitiva necessária para sua realização. Esse nível é usado quando o objetivo é analisar as respostas emitidas pelos estudantes. Subdividimos a classificação da TS em duas categorias: superficial e profunda. A AS é formada pelos níveis uniestructural e multiestructural; por sua vez, a AP é formada pelos níveis relacional e abstrato estendido (BIGGS; COLLIS, 1982).

A análise e classificação dos itens do TPU2022-1 foram realizadas usando critérios da TS, utilizando uma adaptação do modelo de classificação da complexidade dos itens de uma avaliação proposto por Ceia (2002). Por esse modelo, o item de uma avaliação é analisado a partir da resposta esperada para a questão, considerando três parâmetros: quantidade de conhecimento necessária para encontrar a resolução, operações mentais envolvidas para a

solução dos problemas e complexidade da resposta solicitada (CEIA, 2002). O quadro 2 apresenta o sistema de categorização proposto para este estudo, que foi adaptado dos trabalhos de Ceia (2002) e Pereira (2019).

Quadro 2: Sistema de categorização dos itens do TPU2022-1.

Categoria SOLO da questão	Conhecimento abordado		Procedimentos cognitivos envolvidos
	Número de tópicos utilizados	Relação dos tópicos	
Abstrato (SOLO 5)	Dois ou mais tópicos	Necessita de associação entres os tópicos	indução e/ou dedução; exige identificação de informação relevante, elaboração de hipóteses e generalizações não comumente discutidas no curso médico
Relacional (SOLO 4)			indução e/ou dedução; exige identificação de informações relevantes frequentemente discutidas no curso médico
Multiestrutural (SOLO 3)		Utilizados isoladamente	indução e/ou dedução; exige identificação de informações relevantes frequentemente discutidas no curso médico
Uniestrutural (SOLO 2)	Um único tópico	Não se aplica	indução e/ou dedução; exige identificação de informação relevante frequentemente discutida no curso médico

Fonte: Adaptado de Ceia (2002) e Pereira (2019).

A partir das informações do quadro 2, foi elaborado a Ficha de Categorização do Item (FCI) (APÊNDICE C), onde, para cada item do TPU2022-1, foram transcritos o texto-base, o enunciado, seguido das alternativas e o gabarito da questão. Ainda constam na FCI a justificativa da resposta e espaços para análise do conteúdo do item, dos procedimentos cognitivos necessários à sua resolução e para sua categorização. Uma lista de verbos frequentemente associada a cada categoria da TS foi anexada à FCI para que os avaliadores possam consultar e, assim, melhor ajustar seus pareceres.

A Coordenação do Núcleo de Avaliação do Curso de Medicina da UNIFENAS de Belo Horizonte disponibilizou os comentários pedagógicos para a resolução de cada item do TPU2022-1, o que facilitou a análise da complexidade cognitiva e a classificação de acordo com a TS.

Foram selecionados 33,33% dos itens do TPU BH 2022 1, que correspondem a 40 itens, utilizando-se uma amostragem probabilística sistemática para que um segundo avaliador, no

caso, o orientador desse estudo, realizasse uma análise independente com o intuito de identificar algum viés de categorização.

As análises do pesquisador e do orientador foram confrontadas e verificou-se uma alta taxa de concordância: 95%. Ajustes foram então realizados, permitindo uma calibração da categorização por parte do pesquisador. Os avaliadores, pesquisador e orientador, são médicos e receberam capacitação e treinamento sobre os pressupostos da TS. Os demais itens do TPU2022-1 foram analisados e classificados somente pelo pesquisador.

4.6 Análise Psicométrica pela Teoria Clássica dos Testes

Utilizamos no nosso estudo a TCT e calculamos para cada IME do TPU2022-1 o IDF e o ID. Os índices foram calculados com base no resultado das provas dos alunos do último ano do curso (11º e 12º períodos).

De acordo com a TCT, o IDF de um item é estimado pelo percentual de acertos obtidos pelos estudantes em relação ao total de respondentes participantes daquela avaliação. Um item tem maior nível de dificuldade quando menos pessoas o acertam (MOL, 2019). O IDF varia de zero a um, sendo que, quanto mais próximo de zero, o item é considerado mais difícil, e, quanto mais próximo de um, é considerado mais fácil. A tabela 2 explicita a classificação do IDF.

Tabela 2: Distribuição e classificação do grau de dificuldade dos itens em uma avaliação pela Teoria Clássica dos Testes.

Distribuição esperada (%)	Classificação do item em relação ao índice de dificuldade	
10%	Superior a 0.9	Muito fáceis
20%	De 0.7 a 0.9	Fáceis
40%	De 0.3 a 0.7	Medianos
20%	De 0.1 a 0.3	Difíceis
10%	Até 0.1	Muito difíceis

Fonte: Vilarinho (2015, p. 27).

Como consta na tabela acima, espera-se um equilíbrio entre os índices de dificuldade dos itens de uma avaliação, devendo haver uma porcentagem maior de itens medianos, evitando uma proporção significativa de itens muito difíceis ($< 0,1$) e muito fáceis ($> 0,9$) (FIGUEIREDO, 2008).

O ID relaciona-se com a coerência do score do item com o do teste. Esse índice é calculado pela diferença entre o percentual de acerto de um determinado item dos 27% de estudantes que

tiveram melhor desempenho no teste, dos 27% que tiveram pior desempenho (FONTANIVE; ELLIOT; KLEIN, 2007). Quanto maior a diferença encontrada, maior será o ID. O ideal é que esse índice seja superior a 0,40 (VILARINHO, 2015). Um ID muito baixo ($< 0,20$) indica que o item não foi capaz de separar devidamente os alunos de melhor e pior desempenho, demonstrando que alunos de melhor desempenho obtiveram uma porcentagem de acerto menor do que os de pior desempenho naquele item (FONTANIVE; ELLIOT; KLEIN, 2007).

A Tabela 3 apresenta a classificação dos itens de acordo com o poder de discriminação pela TCT.

Tabela 3: Classificação do índice de discriminação pela Teoria Clássica dos Testes

Índice de Discriminação	Classificação
Discriminação ≥ 0.40	Item bom
$0.30 \leq$ Discriminação < 0.40	Item bom, mas sujeito a aprimoramento
$0.20 \leq$ Discriminação < 0.30	Item marginal, sujeito a reelaboração
Discriminação < 0.20	Item deficiente, deve ser rejeitado

Fonte: Vilarinho (2015, p. 28).

4.7 Análise estatística

Com o objetivo de investigar se houve diferença significativa nas médias das medidas dos parâmetros IDF e ID quanto aos níveis do SOLO, a Análise de Variância com 1 fator (*OneWay*) foi aplicada aos dados. O Teste de Igualdade de Variância de Levene foi utilizado para investigar se as variâncias entre as categorias eram estatisticamente diferentes. Além disso, o tamanho de efeito (“ETA Parcial al Quadrado”) foi calculado. Para investigar se ocorreu diferença significativa nas médias das medidas dos parâmetros IDF e ID quanto às categorias SOLO que representam o aprendizado superficial e o aprendizado profundo (2 ou 3×4), o teste T de Student para Amostras Independentes foi aplicado aos dados. O Teste de Igualdade de Variância de Levene foi novamente utilizado para investigar se as variâncias entre as categorias eram estatisticamente diferentes. Além disso, o Tamanho de Efeito (d de Cohen) foi calculado.

4.8 Aspectos éticos

A presente pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (CEP) da UNIFENAS, sob o CAAE n. 56009222.9.0000.5143, parecer n. 5.379.183 (ANEXO A). Foi solicitada a dispensa do Termo de Consentimento Livre e Esclarecido (TCLE) e encaminhados o Termo de Compromisso de Utilização de Dados (TCUD) e o Termo de Anuência (TA).

5 RESULTADOS

5.1 *Análise da complexidade cognitiva dos itens*

Encontramos um equilíbrio entre as aprendizagens superficial e profunda nos itens do TPU2022-1. No total desses itens, cerca de 40,8% foram classificados como AS e 59,2% como AP, como mostra a tabela 4. Quando analisamos por áreas específicas do conhecimento, verificamos um equilíbrio entre as aprendizagens superficial e profunda nas áreas de cirurgia, ginecologia-obstetrícia e pediatria. Na área de ginecologia-obstetrícia, verificamos que 42,7% dos itens relacionavam-se à AS e 58,3% à AP. Na área de cirurgia, cerca de 30% dos itens pertenciam à AS e cerca de 70% à AP. Na área de pediatria, verificamos que cerca de 37,5% dos itens relacionavam-se à AS e 62,5% à AP.

Nas áreas de clínica médica e saúde coletiva, não encontramos esse equilíbrio. Na área de clínica médica, constatamos que 87,5% dos itens analisados estavam relacionados à AP e 12,5% à AS. Já na área de saúde coletiva, verificamos que cerca de 83,3% dos itens relacionavam-se à AS e 16,7% à AP. A tabela 4 divide os itens do TPU2022-1 entre AS e AP.

Tabela 4: Distribuição dos itens do TPU2022-1 entre aprendizagem superficial e aprendizagem profunda.

Área/Solo	Aprendizagem Superficial	Aprendizagem Profunda
Cirurgia	7 (29,2%)	17 (70,8%)
Clínica Médica	3 (12,5%)	21 (87,5%)
Gineco/Obst.	10 (41,7%)	14 (58,3%)
Pediatria	9 (37,5%)	15 (62,5%)
Saúde Coletiva	20 (83,3%)	4 (16,7%)
Total	49 (40,8%)	71 (59,2%)

Não encontramos em nenhum item do TPU2022-1 relacionado ao nível abstrato estendido, nível de maior complexidade cognitiva da TS. A tabela 5 apresenta o resultado da análise do TPU2022-1, de acordo com os níveis de complexidade cognitiva da TS, estando dividida pelas grandes áreas do conhecimento: cirurgia, clínica médica, ginecologia-obstetrícia, pediatria e saúde coletiva.

Tabela 5: Distribuição dos itens do TPU2022-1 pelos níveis da TS.

Área/Solo	Uniestrutural	Multiestrutural	Relacional	Abstrato
Cirurgia	3 (12,5%)	4 (16,7%)	17 (70,8%)	0
Clínica	2 (8,3%)	1 (4,2%)	21 (87,5%)	0
Gineco/Obst.	7 (29,2%)	3 (12,5%)	14 (58,3%)	0
Pediatria	7 (29,2%)	2 (8,3%)	15 (62,5%)	0
SC	18 (75,0%)	2 (8,3%)	4 (16,7%)	0
Total	37 (30,8%)	12 (10%)	71 (59,1%)	0

Apresentaremos, a seguir, exemplos de itens pertencentes a cada nível de complexidade da TS encontrados na análise do TPU2022-1.

Uniestrutural

Item 61 - Mulher, 25 anos, solteira, com vida sexual ativa, procura a Unidade Básica de Saúde (UBS) com queixas ginecológicas de atraso menstrual e sangramento vaginal. A pré-consulta foi realizada pela equipe de enfermagem. A paciente, após os devidos esclarecimentos sobre o seu atendimento pelo médico-professor, pede para não ser avaliada na presença dos alunos. A conduta MAIS apropriada do ponto de vista da relação médico-professor-aluno-paciente é

Alternativas

- a. realizar o atendimento com a presença dos alunos.
- b. solicitar ultrassonografia pélvica, sem realizar o exame clínico completo.
- c. respeitar, integralmente, o desejo da paciente.
- d. encaminhar para outro serviço.

O gabarito é a letra “c”.

Esse item foi classificado como uniestrutural, pois exigiu uma retenção de um detalhe factual, através de memorização (direito previsto pelo Código de Ética, que veta ao médico a prática da medicina, no exercício da docência, sem o consentimento do paciente ou de seu representante legal), não exigindo relações de tópicos para sua realização.

Multiestrutural

Item 91 - Mulher, 44 anos, apresenta nódulo tireoidiano de cerca de um ano de evolução. Diante da suspeita de neoplasia, é solicitada punção aspirativa por agulha fina (PAAF). Podemos afirmar que:

Alternativas:

- a. o carcinoma anaplásico tem crescimento lento, melhor prognóstico e baixa malignidade.
- b. o carcinoma medular tem origem nas células C, produz tireoglobulina e responde a supressão com TSH.
- c. o carcinoma folicular tem origem nas células foliculares, predominância no gênero masculino e o tratamento é a radioterapia.
- d. o câncer papilífero está 90% relacionado à radiação, muitas vezes é multifocal e com o PAAF pode ser encontrado corpos psamomatosos.

O gabarito é a letra “d”.

Esse item foi classificado como multiestructural porque exigiu uma retenção de detalhes factuais, através de memorização (conhecimento das alterações histológicas, epidemiologia e terapêutica das neoplasias tireoidianas). Não houve necessidade de relação de tópicos para a realização desse item.

Relacional

Item 118 - Menina, 13 anos, é atendida em UBS com queixa de cefaleia, tosse seca, dores no corpo e febre de 38°C, de início há dez dias, além de uma discreta falta de ar que se iniciou há um dia. Nega alergias ou doenças prévias. Informa que seu pai está com tosse seca persistente. Foi medicada em casa com dipirona. Exame físico: Tax: 37,9°C, FC: 80 bpm, FR: 24 irpm, discretas crepitações e raros sibilos em hemitórax direito. Quais são, respectivamente, o diagnóstico mais provável e a conduta adequada ao caso?

Alternativas:

- a. Gripe pelo vírus influenza A/H1N1; prescrição de sintomáticos.
- b. Pneumonia atípica; prescrição de azitromicina e antitérmicos e reavaliação em 48 horas.
- c. Crise de asma; inalação com broncodilatador beta2 de curta ação e corticoide oral por 5 dias.
- d. Pneumonia bacteriana; azitromicina, sintomáticos e reavaliação em 48 horas.

Gabarito da questão: “b”.

Esse item foi classificado como relacional porque exigiu a integração dos dados da história clínica e do exame físico para argumentar a mais provável hipótese diagnóstica e o planejamento da conduta mais adequada (raciocínio clínico). O item exigiu aplicação de conhecimentos, demandando a relação de tópicos e conceitos para sua resolução.

Verificamos que os itens do TPU2022-1 compostos por casos clínicos que exigiam habilidades de resolução de problemas, cognição e integração de conhecimento propiciaram maior exploração da AP. A tabela 6 apresenta a estratégia empregada na formulação dos itens do TPU2022-1, baseada na presença ou ausência de casos clínicos. A maioria dos itens (84,2%) apresentava um texto-base com a presença de um caso clínico. Verifica-se a necessidade da relação de tópicos para resolução de casos clínicos em grande parte dos itens das áreas de conhecimento, com exceção da área de saúde coletiva, na qual uma considerável parcela dos itens (54,2%) não continha casos clínicos, e quando um caso clínico estava presente, o que ocorreu em 11 itens (45,8%), somente em dois foi exigida a relação de tópicos para sua resolução. A tabela 7 correlaciona a presença ou ausência de casos clínicos com a classificação SOLO.

Tabela 6: Estratégia de formulação dos itens do TPU2022-1, indicando a presença ou ausência de casos clínicos.

Área/Item SOLO	Ausência de caso clínico	Presença de caso clínico
Cirurgia	1	23
Clínica Médica	0	24
Gineco/Obst.	3	21
Pediatria	2	22
Saúde Coletiva	13	11
Total	19	101

Tabela 7: Estratégia de formulação dos itens do TPU2022-1, indicando a presença ou ausência de casos clínicos e a classificação SOLO.

Área/Item SOLO	ACC2	ACC3	ACC4	PCC2	PCC3	PCC4
Cirurgia	0	1	0	3	3	17
Clínica Médica	0	0	0	2	1	21
Gineco/Obst.	2	1	0	5	2	14
Pediatria	2	0	0	5	2	15
Saúde Coletiva	10	1	2	8	1	2
Total	14	3	2	23	9	69

ACC2: ausência de caso clínico SOLO 2 - ACC3: ausência de caso clínico SOLO 3 - ACC4: ausência de caso clínico SOLO 4 - PCC2: presença de caso clínico SOLO 2 - PCC3: presença de caso clínico SOLO 3 - PCC4: presença de caso clínico SOLO 4

5.2 *Análise Psicométrica pela Teoria Clássica dos Testes*

Incluímos na análise psicométrica do nosso estudo somente os alunos do último ano do curso (11º e 12º períodos).

5.2.1 Índice de Dificuldade

Verificamos que 86,6% dos itens do TPU2022-1 encontraram-se dentro dos índices fácil, mediano e difícil, sendo que 50,8% encontraram-se dentro do IDF mediano. A proporção de itens muito difíceis e muito fáceis foi de 13,3%. A Tabela 8 apresenta os resultados de acordo com o IDF.

Tabela 8: Índice de Dificuldade TPU2022-1.

% de acerto	N	%
≤ 10,0% (Muito Fácil)	1	0,8
De 10,1 a 30,0% (Fácil)	13	10,8
De 30,1 a 70,0% (Mediana)	61	50,9
De 70,1 a 90,0% (Difícil)	30	25,0
> 90,0% (Muito difícil)	15	12,5
Total	120	100,0

5.2.2 Índice de Discriminação

Verificamos na nossa análise que 79,2% dos itens apresentaram um ID menor que 20%. Cerca de 2% dos itens apresentaram um ID maior que 40%. A Tabela 9 apresenta os resultados de acordo com o índice de Discriminação.

Tabela 9: Índice de Discriminação TPU2022-1.

Discriminação	n	%
< 20,0%	95	79,1
De 20 a 29,9%	15	12,5
De 30 a 39,9%	8	6,7
≥ 40,0%	2	1,7
Total	120	100,0

5.3 Correlação entre os Índices de Dificuldade e de Discriminação com a Classificação SOLO

A tabela 10 e os gráficos 1 e 2 mostram que não verificamos diferença estatisticamente significativa entre as três categorias SOLO quanto às médias do IDF ($p = 0,821$; $F_{2, 117} = 0,197$) e o ID ($p = 0,484$; $F_{2, 117} = 0,197$), ou seja, a igualdade entre as três categorias do SOLO não é rejeitada ($p \geq 0,05$). Ressalta-se que o *Tamanho de efeito* calculado é considerado pequeno, portanto, corroborando com a não significância estatística do teste ($p > 0,05$). Cabe ressaltar que não foi observada diferença estatisticamente significativa entre as variâncias (*Levene* $\rightarrow p \gg 0,05$). Logo, não há necessidade da aplicação do teste de *Welch*.

Tabela 10: Análises descritiva e comparativa entre as categorias do SOLO quanto aos parâmetros de avaliação do TCT (Dificuldade e Discriminação).

SOLO	Medidas descritivas		P
	N	Média \pm d. p.	
Dificuldade			
2	37	60,9 \pm 22,8	0,821 $F_{2, 117} = 0,197$
3	12	62,1 \pm 19,4	
4	71	58,5 \pm 24,4	
Geral	120	59,6 \pm 23,3	
Discriminação			
2	37	15,0 \pm 10,6	0,484 $F_{2, 117} = 0,731$
3	12	11,3 \pm 7,2	
4	71	13,4 \pm 9,6	
Geral	120	13,7 \pm 9,7	

Base de dados: 120 questões (SOLO 2 \rightarrow 37 casos, SOLO 3 \rightarrow 12 casos e SOLO 4 \rightarrow 71 casos)

Nota: **d.p.** \rightarrow Desvio-padrão.

p \rightarrow Probabilidade de significância da Análise de Variância com 1 fator (*OneWay*).

F \rightarrow Estatística da Análise de Variância com 1 fator (*OneWay*).

- **Teste de Levene para igualdade de variância:** $p \gg 0,05$. Para ambas as variáveis.

- **Tamanho de efeito (η^2):** $\eta^2 \leq 0,01 \rightarrow$ Tamanho de efeito Pequeno (ambas as variáveis).

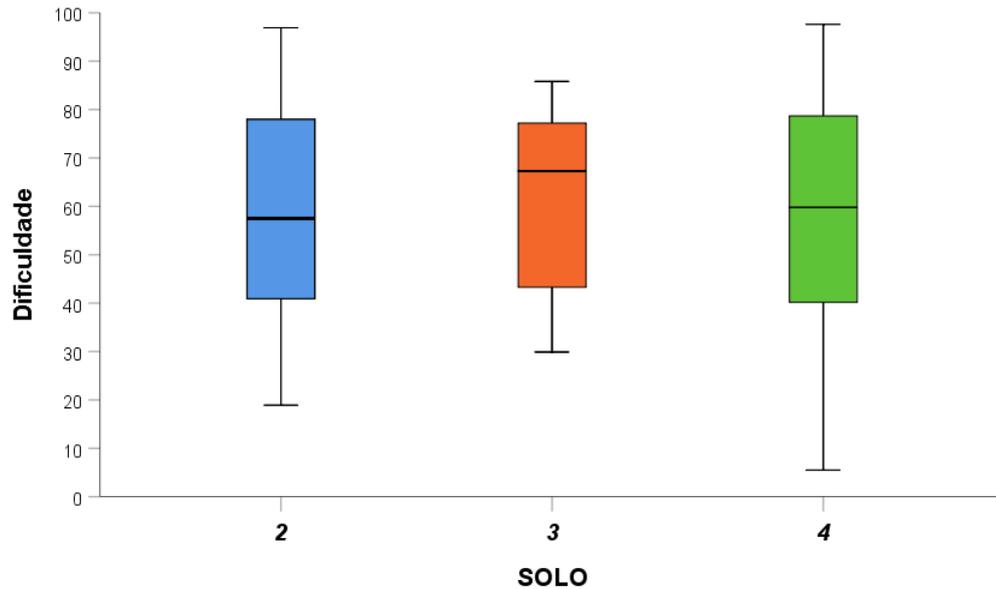


Gráfico 1: BOXPLOT do parâmetro IDF, por SOLO.

Notas: **Base de Dados:** 120 questões (SOLO 2 → 37 casos, SOLO 3 → 12 casos e SOLO 4 → 71 casos).

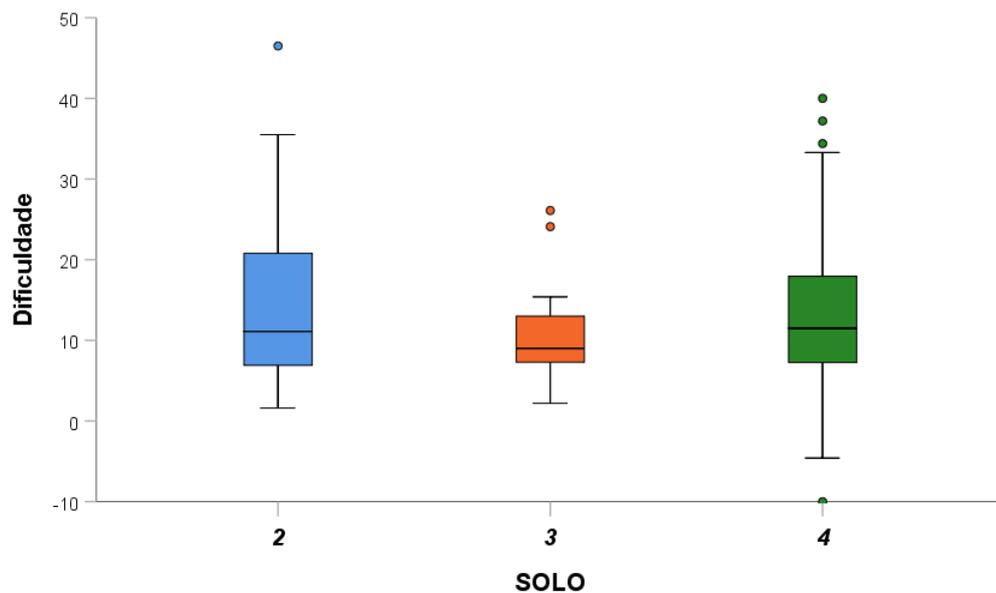


Gráfico 2: BOXPLOT do parâmetro ID, por SOLO

Notas: **Base de Dados:** 120 questões (SOLO 2 → 37 casos, SOLO 3 → 12 casos e SOLO 4 → 71 casos).

A tabela 11 e os gráficos 3 e 4 mostram que também não verificamos diferença estatisticamente significativa entre as categorias SOLO que representam o aprendizado superficial e o aprendizado profundo quanto às médias do IDF ($p = 0,544$; $t_{118} = 0,609$) e o ID ($p = 0,684$; $t_{118} = 0,408$).

Tabela 11: Análises descritiva e comparativa entre duas categorias do SOLO quanto aos parâmetros de avaliação do TCT (Dificuldade e Discriminação).

SOLO	Medidas descritivas		P
	N	Média ± d. p.	
Dificuldade			
2 ou 3	49	61,2 ± 21,8	0,544 $t_{118} = 0,609$
4	71	58,5 ± 24,4	
Geral	120	59,6 ± 23,3	
Discriminação			
2 ou 3	49	14,1 ± 9,9	0,684 $t_{118} = 0,408$
4	71	13,4 ± 9,6	
Geral	120	13,7 ± 9,7	

Base de dados: 120 questões (SOLO 2 → 37 casos, SOLO 3 → 12 casos e SOLO 4 → 71 casos)

Nota: **d.p.** → Desvio-padrão

p → Probabilidade de significância da Análise de Variância com 1 fator (*OneWay*).

F → Estatística da Análise de Variância com 1 fator (*OneWay*).

- **Teste de Levene para igualdade de variância:** $p \gg 0,05$. Para ambas as variáveis.

- **Tamanho de efeito (*d de Cohen*):** $d \leq 0,12$ → Tamanho de efeito Pequeno (ambas as variáveis).

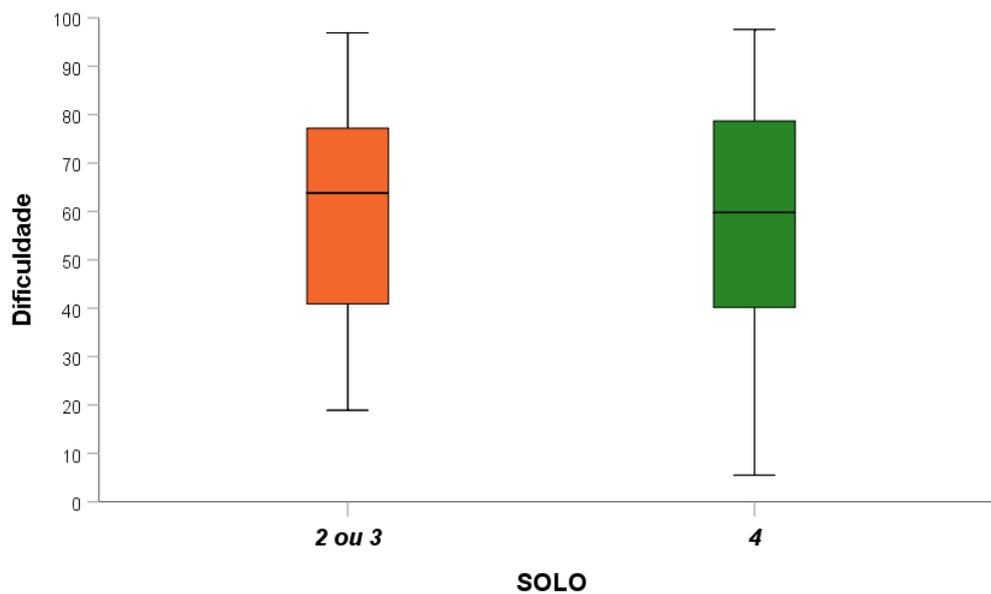


Gráfico 3: BOXPLOT do parâmetro “Dificuldade”, por SOLO (2 categorias)

Nota: **Base de dados:** 120 questões (SOLO 2 ou 3 → 49 casos e SOLO 4 → 71 casos)

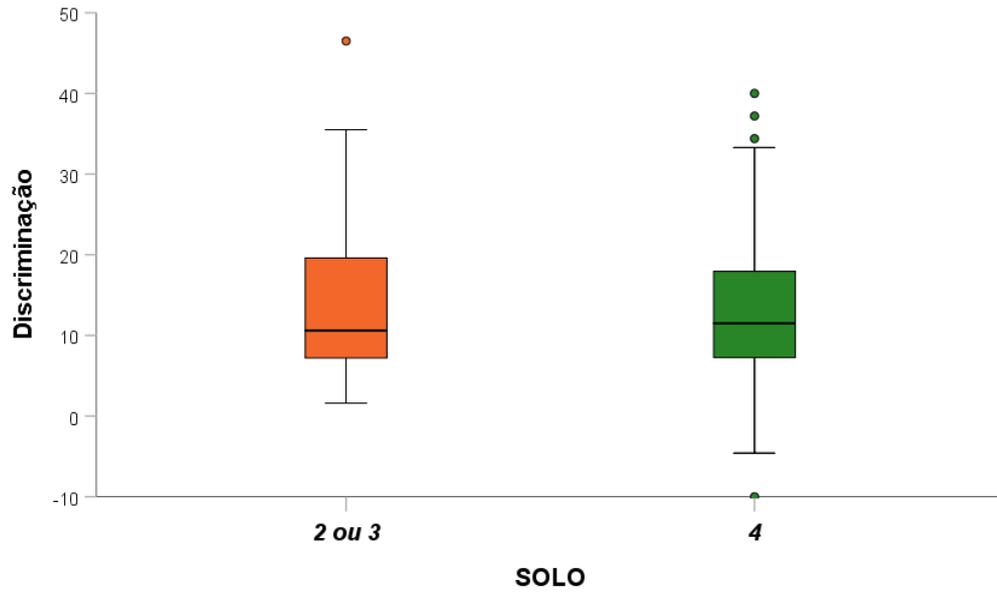


Gráfico 4: BOXPLOT do parâmetro “Discriminação”, por SOLO (2 categorias)

Nota: **Base de dados:** 120 questões (SOLO 2 ou 3 → 49 casos e SOLO 4 → 71 casos).

6 DISCUSSÃO

A TS é uma ferramenta que possibilita a adequação dos métodos de avaliação ao programa do currículo das escolas médicas, além de facilitar e orientar o princípio da progressividade do processo de ensino e aprendizagem. Também permite realizar atividades avaliativas relacionadas ao nível de aprendizagem desejável aos discentes. (KORKMAZ; UNSAL, 2017). Os níveis uniestructural e multiestructural da TS se relacionam a um crescimento quantitativo e progressivo de conhecimentos. Já os níveis relacional e abstrato estendido associam-se a um crescimento qualitativo, responsável pela organização e estruturação do conhecimento (relacional) que podem levar à teorização e generalização (abstrato estendido) (SOOBARD; RANNIKMÄE, 2015). O pensamento profundo, relacionado à fase qualitativa da TS, é aquele capaz de raciocinar, explicar e generalizar. É aquele capaz de solucionar problemas (CANIGLIA; MEADOWS, 2018).

Constatamos um equilíbrio entre a AS e a AP no TPU2022-1, como mostra a tabela 4. Quando dividimos por áreas de conhecimento, verificamos um desequilíbrio nas áreas de clínica médica e de saúde coletiva. Esse desequilíbrio pode ser explicado pela estratégia adotada na formulação dos itens. Na área de clínica médica, todos os itens foram compostos de casos clínicos, tendo a maior parte desses (87,5%) a necessidade da relação de tópicos para sua resolução. Na área de saúde coletiva, uma grande parte dos itens não continha casos clínicos (54,2%) e, quando um caso clínico se encontrava no texto base, o que ocorreu em 11 itens (45,8%), somente em dois havia a necessidade de relação de tópicos e conceitos para sua resolução.

Pesquisadores defendem que um equilíbrio na distribuição dos níveis de pensamento, relacionado à TS, contribui para uma melhor aprendizagem. Dessa forma, ganha grande importância uma distribuição balanceada nos itens de uma avaliação de acordo com as taxonomias de aprendizagem (KORKMAZ; UNSAL, 2017). Apesar da importância atribuída a esse equilíbrio entre a AS e a AP nos processos de avaliação, é de se esperar que os egressos do curso médico estejam aptos para a resolução de problemas complexos. A TS auxilia na escolha e na construção de itens de avaliação adequados para a verificação de aquisição de atributos esperados de um graduado e que fomentem a análise ao invés de uma simples memorização (HADIE, 2018). A expectativa é que os alunos dos últimos períodos estejam mais habilitados para a resolução de atividades complexas, como solução de problemas e tomadas de decisões (SOOBARD; RANNIKMÄE, 2015).

Os itens classificados como aprendizagem profunda do TPU2022-1 continham, na grande maioria das vezes, um caso clínico no texto-base e o enunciado aferia habilidades de resolução de problemas, cognição e integração de conhecimento, fomentando um raciocínio clínico. Por sua vez, o raciocínio clínico demanda uma base de conhecimento, propiciando aos estudantes a geração de hipóteses, estabelecimento de diagnósticos e apresentação de uma conduta para a resolução de problemas clínicos. É uma habilidade fundamental ao egresso do ensino médico (PEIXOTO *et al.*, 2018; THAMPY; WILLERT; RAMANI, 2019). O uso de resoluções de problemas clínicos autênticos é uma medida eficaz de avaliação do raciocínio clínico (VAN DER VLEUTEN; SCHUWIRTH, 2019).

A construção do pensamento de ordem superior é imprescindível na formação e no exercício da medicina, tendo o raciocínio clínico e o pensamento crítico, objetivos primordiais no decorrer do curso (ARAGÃO; ALMEIDA, 2017). A AP é relacionada a uma melhor qualidade e efetividade no ensino médico e está mais associada à capacidade de os estudantes se atualizarem após a sua formação básica (ROSSI *et al.*, 2021).

O método de questões de múltipla escolha é muito utilizado nos processos de avaliação das escolas médicas. As avaliações baseadas em IME, quando bem concebidas, possuem a capacidade de avaliar os alunos nos níveis superiores do conhecimento (VEGI *et al.*, 2022). No ensino médico, a inserção de casos clínicos nos testes de múltipla escolha aprimora a qualidade desse processo de avaliação para a mensuração do raciocínio clínico (MODI *et al.*, 2015). Verificamos isso no TPU2022-1, onde muitos itens exigiram um raciocínio mais complexo através da resolução de casos clínicos. Os IMEs que demandam a aplicação de conhecimento, exigindo a relação de tópicos e conceitos para sua resolução, têm o potencial para a avaliação da aprendizagem de nível superior (SCULLY, 2017).

Os IMEs baseados em casos clínicos são práticos, seguros e facilmente alinhados à ABP (VUMA; SA, 2017). A ABP favorece a AP e autodirigida (VAN DER VLEUTEN; SCHUWIRTH, 2019). Ainda quanto aos IMEs, eles são muito utilizados nos concursos de pós-graduação, como residência médica, e, quando bem formulados, permitem que os discentes sejam avaliados para a aplicação e a análise, o que possibilita a aferição da AP (VEGI *et al.*, 2022).

Scully (2017) cita estratégias que orientam a construção de IME para avaliar o pensamento de ordem superior, a AP. A primeira estratégia é o uso de verbos associados ao nível cognitivo

desejável, porém deve-se ter muito cuidado, pois alguns verbos podem pertencer a mais de um nível taxonômico, dependendo muito do contexto do item. Com objetivo de alcançar um nível taxonômico de ordem superior, pode-se usar o artifício de substituir, no enunciado, o verbo por frases como: “selecione o melhor plano terapêutico”, “indique a melhor hipótese diagnóstica”. A segunda estratégia citada pela autora é a de itens invertidos, na qual, em vez de constar nas alternativas de resposta a explicação de um conceito solicitado no enunciado, é apresentado no texto-base o esclarecimento e nas alternativas, os temas. Isso demanda dos alunos o conhecimento dos conceitos dos distratores. A terceira estratégia é o uso de distratores de alta qualidade, que requerem um nível maior de discriminação para a resposta. Itens com distratores incabíveis não avaliam a AP. A autora denomina a quarta estratégia de “uso de múltiplos neurônios”, pois, para sua resolução, demanda dos discentes a aplicação de conhecimento com relação de tópicos. Utiliza-se muito nessa estratégia o uso de cenários, como casos clínicos, onde há a necessidade de se basear na análise e relação de vários conteúdos para a resolução do item.

Cito como exemplo de estratégia de itens invertidos a questão de número 25 do TPU2022-1, em que constam nas alternativas de resposta os possíveis diagnósticos, e no texto-base, os tópicos a serem relacionados. Essa técnica exige o conhecimento dos distratores.

Item 25 - Uma jovem de 16 anos vem à consulta por amenorreia primária. Tem 1,60 m de altura, 55 kg e IMC de 21,4. Apresenta desenvolvimento sexual secundário completo e hímen íntegro. Os testes de progesterona e o de progesterona mais estrógeno são negativos. A PRINCIPAL hipótese diagnóstica nesse caso é a de:

Alternativas:

- a. síndrome de Turner.
- b. síndrome de Mayer-Rokitansky-Kuster-Hauser.
- c. síndrome de Morris.
- d. hipotireoidismo.

Gabarito da questão: “b”

Observamos no item 88 a terceira estratégia de Scully, pois todos os distratores são plausíveis, não havendo nas alternativas um distrator incabível com o quadro clínico. Verifica-se também nessa questão a utilização da técnica dos itens invertidos.

Item 88 - Paciente do sexo feminino, 38 anos, procura UBS com queixa de dor anal exacerbada pela defecação e que persiste de alguns minutos até uma hora. Após o exame físico, o médico assistente conclui tratar-se de infecção das criptas com propagação para o tecido subcutâneo e presença de ulceração na comissura posterior. Diante da queixa da paciente e do exame local do médico assistente, o diagnóstico MAIS PROVÁVEL é o de:

Alternativas:

- a. fístula anal.
- b. fissura anal.
- c. plicoma anal.
- d. hemorroida.

Gabarito da questão: “b”

Como exemplo de estratégia de uso de múltiplos neurônios, citada por Scully (2017), cito a questão abaixo, onde houve a necessidade da relação e da integração de conhecimentos, fomentando um raciocínio clínico, para a apresentação da melhor conduta para o caso.

Item 73 - Homem, 45 anos, politraumatizado, foi submetido à laparotomia exploradora por ferimento penetrante no abdome produzido por projétil de arma de fogo. Durante a cirurgia, foram corrigidos vários danos às estruturas intra-abdominais. Ainda na sala de recuperação pós-anestésica, o paciente passou a apresentar distensão abdominal, oligúria, acidose metabólica, aumento da pressão venosa central, sinais de hipertensão intracraniana, sinais de insuficiência respiratória com hipercapnia e hipóxia, diminuição da complacência pulmonar e da capacidade funcional residual. Considerando o quadro apresentado pelo paciente, a MELHOR conduta a ser adotada nesse momento é

Alternativas:

- a. reprogramar a ventilação mecânica, reduzindo a frequência respiratória.
- b. retornar ao bloco cirúrgico para fechamento alternativo da parede abdominal.
- c. reduzir a quantidade de relaxante muscular administrada ao paciente.
- d. realizar hemotransfusão.

Gabarito da questão: “b”

Vemos também no item 73 a primeira estratégia citada por Scully (2017), onde há a substituição, no enunciado, de um verbo de comando pela frase que seleciona o melhor plano terapêutico.

Javaeed (2018) analisou a capacidade dos IMEs para avaliar o pensamento de ordem superior na educação médica e concluiu que um IME, quando bem estruturado, tem a capacidade de avaliar a aprendizagem profunda. Os IMEs bem estruturados foram superiores a questões dissertativas modificadas, que são compostas por cenários clínicos curtos seguidos por uma série de questões com um formato estruturado para pontuação, a fim de testar habilidades cognitivas superiores em estudantes de medicina (KHAN; ALJARALLAH, 2011).

Casiraghi e Aragão (2020) avaliaram e classificaram, pela TB, 429 itens de provas realizadas em um curso de medicina do estado do Rio de Janeiro durante o ano de 2013, englobando várias áreas de conhecimento (bases humanísticas, ciências básicas, saúde coletiva, clínica médica, clínica cirúrgica, pediatria e ginecologia/obstetrícia). Verificaram que 46,62% das questões relacionavam-se a conhecimento; 14,4%, compreensão; 15,8%, aplicação; 5,6%, análise; 14,2%, síntese e 3,2%, julgamento. Do total dos 429 itens, 99 (23,1%) foram considerados de raciocínio clínico e 330 (76,9%) de raciocínio teórico.

Villela *et al.* (2022) verificaram a adequação e a qualidade de IMEs do TP realizado pelos estudantes da Faculdade de Medicina de Ribeirão Preto/USP, no período de 2013 a 2018. Utilizaram sete indicadores de boas práticas, sendo um deles a aplicação do conhecimento ou interpretação de dados em vez de apenas memorização de conteúdo. A grande maioria (97,5%) dos itens foi relacionada à aplicação do conhecimento, pertencente a um nível elevado do domínio cognitivo.

Não encontramos nenhum item do TPU2022-1 relacionado ao nível abstrato estendido, nível de maior complexidade cognitiva da TS. Esse fato não nos surpreendeu, pois os IMEs trazem uma certa limitação para a demonstração de níveis abstratos avançados de aprendizado (SPRECHER, 2019). Scully (2017) menciona a capacidade dos IME avaliarem o pensamento de ordem superior, porém cita a limitação destes para a avaliação dos níveis taxonômicos mais elevados. Além disso, o nível abstrato estendido é mais associado aos programas de pós-graduação. Brabrand e Dahl (2009) justificam que um conhecimento mais aprofundado e qualificado, atingindo os níveis mais elevados de complexidade cognitiva da TS, muitas vezes somente será alcançado na pós-graduação. Outras estratégias de avaliação, como os itens

discursivos, são apropriadas para demonstrar esse nível, no entanto seria impraticável a realização de um exame com esse tipo de item, no qual o objetivo é avaliar um grande número de alunos, com um conteúdo muito abrangente (KIM *et al.*, 2012).

Outra importante aplicabilidade da TS é fornecer ao aluno dados sobre seu nível de pensamento cognitivo. Isso se torna muito importante nesse momento de mudança de paradigma do Ensino Superior, no qual se propõe uma aprendizagem centrada no aluno, autônoma, baseada na teoria construtivista (JAISWAL, 2019).

A análise psicométrica dos itens de uma avaliação é uma ferramenta simples, porém valiosa para a verificação da qualidade dos IMEs, no intuito da obtenção de meios fidedignos ao processo de ensino e aprendizagem. O IME deve ser avaliado apoiado em bases psicométricas para verificar a validade e a confiabilidade, sendo considerado ideal aquele que possui um médio IDF e um alto ID (GAJJAR *et al.*, 2014). Torna-se muito importante analisar a razão da escolha de uma opção de resposta em detrimento de outras e o motivo pelo qual, muitas vezes, a alternativa mais respondida não é a correta. Para uma resposta a esses questionamentos, é necessária uma associação do IDF com o ID (PITON-GONÇALVES; ALMEIDA, 2018).

Verificamos um satisfatório resultado relacionado ao IDF, com uma porcentagem menor de itens muito fáceis e muito difíceis (13,3%). Encontramos 86,7% dos itens do TPU-BH 2022-1 dentro de uma faixa aceitável para um processo de avaliação.

No presente estudo, a maioria dos IME encontrou-se em um nível insatisfatório em relação ao ID, não proporcionando uma diferenciação dos alunos de melhor e de pior desempenho, o que indica a necessidade da revisão de um número significativo de itens para que esse índice fique melhor calibrado.

Alguns fatores podem afetar o ID de um item, como ambiguidade da pergunta, acentuada dificuldade ou facilidade para a sua resolução, presença de conteúdos não discutidos no currículo, má preparação dos discentes, grau de motivação dos alunos, número de questões, tempo para a resolução e fatores ambientais, como temperatura, ruído e ventilação. O baixo ID não tem obrigatoriamente o significado de um item defeituoso e que deva ser eliminado, mas que haja uma análise crítica de sua qualidade. (BHATTACHERJEE *et al.* 2022; VIANNA, 1987).

Gajjar *et al.* (2014), analisando estudantes em um exame interno de medicina comunitária contendo 50 IMEs, encontraram 24 itens com IDFs considerados bons a excelentes e 15 apresentaram IDs classificados como aceitáveis ($> 0,25$). O ID médio foi de 0,14. Bhattacharjee *et al.* (2022) analisaram uma avaliação interna on-line entre estudantes do sexto semestre de medicina contendo 60 IMEs e encontraram 38,3% com IDFs aceitáveis e 76,6% com IDs insatisfatórios.

Quanto à correlação entre os IDFs e IDs com a classificação SOLO, não identificamos diferença estatisticamente significativa entre as três categorias (uniestrutural, multiestrutural e relacional). Também não verificamos diferença estatisticamente significativa dessa correlação entre as categorias SOLO que representam o aprendizado superficial e o aprendizado profundo (uniestrutural e multiestrutural x relacional). Dessa forma, não houve uma relação direta entre os níveis SOLO e os IDFs e IDs no TPU2022-1.

Kim *et al.* (2012) utilizaram a TB na elaboração de avaliações em um curso de farmácia, referindo que IMEs bem elaborados são capazes de avaliar a AP. Verificaram um melhor ID nos itens relacionados ao pensamento crítico comparados com os que exigiam apenas memorização.

Para um melhor desempenho do TP, torna-se fundamental o preparo do corpo docente para a elaboração dos IMEs. Isso requer um programa de educação continuada docente nas escolas médicas para o desenvolvimento de itens de avaliação de alta qualidade (VANDERBILT; FELDMAN; WOOD, 2013). As taxonomias de aprendizagem podem contribuir de maneira significativa para a construção de avaliações bem planejadas, colaborando para um ambiente eficaz de aprendizagem (HADIE, 2018).

Nos dias de hoje, dispomos de um número grande de técnicas de avaliação. Nenhum método isolado deve ser utilizado nos processos de avaliação das escolas de medicina (SHAH *et al.*, 2012). Para que haja uma satisfatória comprovação do desempenho dos alunos, é necessária uma combinação de técnicas (FOWELL; BLIGH, 1998; KHAN; ALJARALLAH, 2011).

Encontramos na literatura poucos estudos que analisam avaliações externas à TS. Mol (2019) fez uma revisão de estudos brasileiros que utilizam a TS e encontrou 14 trabalhos, dos quais dez são artigos e quatro são dissertações. Nenhuma tese que abordasse a TS foi encontrada. Somente dois estudos foram relacionados com o Ensino Superior e nenhum na área do ensino médico. Ferreira e Rocha (2020) fizeram um levantamento das teses e dissertações defendidas

no Brasil que utilizaram a TS. Encontraram doze trabalhos, a maioria na área de ciências exatas e nenhum relacionado ao ensino médico.

Algumas limitações do presente estudo devem ser consideradas. A primeira se refere ao fato de o nosso estudo ter compreendido apenas uma prova do TP. A análise de um número maior de provas, principalmente de maneira sequencial, poderá trazer maior subsídio para uma melhor interpretação dos dados. Outra limitação deve-se ao fato de que a TS ainda ter sido pouco utilizada no contexto do ensino médico, impossibilitando a correlação de dados e métodos de análise com outros estudos.

Os resultados demonstrados nesse estudo devem ser analisados com muito respeito pela comunidade acadêmica devido à escassez de estudos semelhantes, tanto no âmbito nacional, como internacional, impedindo a realização de comparação dos resultados.

7 CONCLUSÃO

A análise taxonômica do TPU2022-1 nos permitiu observar a qualidade desse instrumento, útil na avaliação do desempenho cognitivo longitudinal dos estudantes durante o curso. Verificamos um equilíbrio entre a AS e a AP quando analisados todos os itens. Entretanto, não observamos esse equilíbrio nas áreas de clínica médica e saúde coletiva. Constatamos uma relação direta entre os níveis de AP da TS e IMEs compostos por casos clínicos, que exigiam habilidades de resolução de problemas, cognição e integração de conhecimento, fomentando um raciocínio clínico. Constatamos uma boa relação para o IDF, mas não para o ID, indicando-se a necessidade de uma análise crítica dos itens. Não verificamos diferença estatisticamente significativa entre as categorias SOLO que representam o aprendizado superficial e o aprendizado profundo quanto às médias do IDF e o ID.

O TP é um importante instrumento de avaliação do ensino, devendo sua realização ser estimulada nos cursos de medicina. O preparo do corpo docente para a realização de seus itens é fundamental. Oficinas para a orientação na elaboração de itens devem ser oferecidas aos professores. Ferramentas taxonômicas, como a TS, podem contribuir de maneira significativa para a realização das atividades avaliativas, adequando as avaliações ao currículo, permitindo a realização de provas relacionadas ao nível de aprendizagem desejável, facilitando, dessa forma, a progressividade do ensino.

Estudos complementares são necessários para ampliar a consistência do uso dessa ferramenta taxonômica no ensino médico.

8 REFERÊNCIAS

- ANDREASSEN, P.; MALLING, B. How are formative assessment methods used in the clinical setting? A qualitative study. **International Journal of Medical Education**, Aarhus, v. 10, p. 208-215, 2019.
- ARAGÃO, J. C. S.; ALMEIDA, L. S. Raciocínio Clínico e Pensamento Crítico: Desenvolvimento na educação médica. **Revista de Estudios e Investigación en Psicología y Educación**, [s. l.], v. Extr., n. 12, p. 12-20, 2017.
- ARIPIN, M. A. *et al.* Unveiling a new taxonomy in education field. **International Journal of Evaluation and Research in Education**, [s. l.], v. 9, n. 3, p. 524-530, 2020.
- BALDIM, Y. L.; VICENTE, C. A. O.; ARCURI, M. B. O teste de progresso sob a visão do discente. **Revista da Faculdade de Medicina de Teresópolis**, [s. l.], v. 2, n. 1, p. 41-54, 2018.
- BHATTACHERJEE, S. *et al.* Evaluation of Multiple-Choice Questions by Item Analysis, from an Online Internal Assessment of 6th Semester Medical Students in a Rural Medical College, West Bengal. **Indian Journal of Community Medicine**, [s. l.], v. 47, n. 1, p. 92-95, 2022.
- BIGGS, J.; COLLIS, K. **Evaluating the quality of learning: the SOLO Taxonomy**. New York: Academic Press, 1982. 296 p.
- BIGGS, J.; TANG, C. **Teaching for Quality Learning at University**. 3. ed. The Society for Research into Higher Education & Open University Press. 2007. 97 p.
- BOLLELA, V. R.; BORGES, M. C.; TROCON, L. E. A. Avaliação Somativa de Habilidades Cognitivas: Experiência Envolvendo Boas Práticas para a Elaboração de Testes de Múltipla Escolha e a Composição de Exames. **Revista Brasileira de Educação Médica**, Brasília, v. 42, n. 4, p. 74-85, 2018.
- BÓRNEA, E. R.; GONÇALVES, A.; PADOVANI, C. R. Avaliando Conhecimento em DST de Graduandos em Medicina segundo a Taxonomia de Bloom. **Revista Brasileira de Educação Médica**, Brasília, v. 38, n. 1, p. 25-30, 2014.
- BRABRAND, C; DAHL, B. Using the SOLO Taxonomy to Analyze Competence Progression of University Science Curricula. **Higher Education**, v. 58, n. 4, p. 531-549, 2009.
- BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. Resolução no 3, de 20 de junho de 2014. Institui as Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina e dá outras providências. **Diário Oficial da União**, Brasília, 23 jun. 2014.
- CANIGLIA, J. C.; MEADOWS, M. An Application of the Solo Taxonomy to Classify Strategies Used by Pre-Service Teachers to Solve “One Question Problems”. **Australian Journal of Teacher Education**, [s. l.], v. 43, n. 9, p. 74-89, 2018.

CASIRAGHII B.; ARAGÃO J. C. S. Metavaliação no Ensino Superior: Raciocínio Clínico em Provas de Medicina. **Estudos em Avaliação Educacional**, São Paulo, v. 31, n. 76, p. 219-230, 2020.

CEIA, M. A taxonomia SOLO e os níveis de Van Hiele. In: ENCONTRO DE INVESTIGAÇÃO EM EDUCAÇÃO MATEMÁTICA, 11., 2002, Coimbra. **Anais...** Coimbra: Sociedade Portuguesa de Ciências da Educação, 2002.

CHAMPLAIN, A. F. C. Setting and maintaining standards in multiple-choice examinations: Guide supplement 37.2 – Viewpoint. **Medical Teacher**, [s. l.], v. 32, p. 436-437, 2010.

DAVIES, G. R.; PROOPS, H., CAROLAN, C. M. The Development and Use of a Multiple-Choice Question (MCQ) Assessment to Foster Deeper Learning: An Exploratory Web-Based Qualitative Investigation. **Journal of Teaching and Learning**, Ontario, v. 14, n. 1, p. 1-12, 2020.

FERRAZ, A. P. C. M.; BELHOT, R. V. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Gestão e Produção**, São Carlos, v. 17, n. 2, p. 421-431, 2010.

FERREIRA, F. F. G.; ROCHA, M. L. P. C. A Taxonomia SOLO nas teses e dissertações defendidas em Programas de Pós-Graduação no Brasil. **Revista de Matemática, Ensino e Cultura**, [s. l.], v. 15, p. 32-46, 2020.

FERRIS, H.; O'FLYNN, D. Assessment in Medical Education; What Are We Trying to Achieve? **International Journal of Higher Education**, Ontario, v. 4, n. 2, p. 139-144, 2015

FIGUEIREDO, V. L. M. *et al.* Propriedades psicométricas dos itens do teste WISC-III, **Psicologia em Estudo**, Maringá, v. 13, n. 3, p. 585-592, 2008.

FONTANIVE, N.; ELLIOT, L. G.; KLEIN, R. Os desafios da apresentação dos resultados da avaliação de sistemas escolares a diferentes públicos. **REICE – Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2, p. 262-273, 2007.

FOWELL, S. L.; BLIGH, J. G. Recent developments in assessing medical students. **Postgraduate Medical Journal**, Liverpool, v. 74, p. 18-24, 1998.

GAJJAR, S. *et al.* Item and Test Analysis to Identify Quality Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. **Indian Journal of Community Medicine**, Gujarat, v. 39, n. 1, p. 17-20, 2014.

GOLD, J. *et al.* Progress testing 2.0: clinical skills meets necessary Science. **Medical Education Online**, Philadelphia (PA), v. 20, p. 1-4, 2015.

HADIE, S. N. H. The Application of Learning Taxonomy in Anatomy Assessment in Medical School. **Education in Medicine Journal**, Kelantan, v. 10, n. 1, p. 13-23, 2018.

HATTIE, J. A. C.; BROWN, G. T. L. **Cognitive processes in asTTle**: the SOLO taxonomy. AsTTle Technical Report. Wellington: Ministry of Education, 2004. p. 41.

ILGUY, M. *et al.* Comparison of Case-Based and Lecture-Based Learning in Dental Education Using the SOLO Taxonomy. **Journal of Dental Education**, [s. l.], v. 78, n. 11, p. 1521-1527, 2014.

IRVINE, J. A comparison of revised Bloom and Marzano's New Taxonomy of Learning. **Research in Higher Education Journal**, Jacksonville (FL), v. 33, p. 1-16, 2017.

IRVINE, J. Marzano's New Taxonomy as a framework for investigating student affect. **Journal of Instructional Pedagogies**, [s. l.], v. 24, p. 2-31, 2017.

JAISWAL, P. Using Constructive Alignment to Foster Teaching Learning Processes. **English Language Teaching**, [s. l.], v. 12, n. 6, p. 10-23, 2019.

JAVAEED, A. Assessment of higher ordered thinking in medical education: multiple choice questions and modified essay questions. **MedEdPublish**, London, v. 7, n. 128, p. 1-13, 2021.

JIMOYIANNIS, A. Using SOLO taxonomy to explore students' mental models of the programming variable and the assignment statement. **Themes in Science & Technology Education**, [S.l.], v. 4, n. 2, p. 53-74, 2011.

KHAN, M. U. Z.; ALJARALLAH, B. M. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. **International Journal of Health Sciences**, Qassim, v. 5, n. 1, p. 39-43, 2011.

KIM, M. *et al.* Incorporation of Bloom's Taxonomy into Multiple-Choice Examination Questions for a Pharmacotherapeutics Course. **American Journal of Pharmaceutical Education**, Arlington (VA), v. 76, n. 6, p. 1-8, 2012.

KNIGHT, C. C.; SUTTON, R. E. Neo-Piagetian Theory and Research: Enhancing Pedagogical Practice for Educators of Adults. **London Review of Education**, London, v. 2, n. 1, 2004.

KORKMAZ, F.; UNSAL, S. Analysis of Attainments and Evaluation Questions in Sociology Curriculum according to the SOLO Taxonomy. **Eurasian Journal of Educational Research**, Ankara, v. 69, p. 75-92, 2017.

LISTER, R. *et al.* Not seeing the forest for the trees: novice programmers and the SOLO taxonomy. In: ANNUAL SIGCSE CONFERENCE ON INNOVATION AND TECHNOLOGY IN COMPUTER SCIENCE EDUCATION (ITICSE '06), 11., 2006, New York (NY). **Proceedings...** Association for Computing Machinery, New York (NY): ACM digital Library, 2006. v. 38, n. 3, p. 118-122.

MODI *et al.* Teaching and Assessing Clinical Reasoning Skills. **Indian Pediatrics**, New Delhi, v. 52, p. 787-794, 2015.

MOL, S. M. **Prova Brasil**: uma análise da complexidade cognitiva de itens de Matemática por meio da Taxonomia SOLO. 2019. 127 f. Dissertação (Mestrado em Educação) - Instituto de Ciências Humanas e Sociais, Universidade Federal de Ouro Preto, Mariana, 2019.

MOL, S. M.; MATOS, A. S. M. Uma análise sobre a taxonomia solo: aplicações na avaliação educacional. **Estudos em Avaliação Educacional**, São Paulo, v. 30, n. 75, p. 722-747, 2019.

O'SHAUGHNESSY, S. M.; JOYCE, P. Summative and Formative Assessment in Medicine: The Experience of an Anaesthesia Trainee. **International Journal of Higher Education**, Ontario, v. 4, n. 2, p. 198-206, 2015.

PEIXOTO, J. M. *et al.* Processos de Desenvolvimento do Raciocínio Clínico em Estudantes de Medicina. **Revista Brasileira de Educação Médica**, Brasília, v. 42, n. 1, p. 75-83, 2018.

PEREIRA, V. C. A. S. **Aplicação da Taxonomia SOLO na análise da qualidade da avaliação**: Validação do método analítico por aplicação aos exames nacionais de Matemática A entre 2006 e 2014. 2019. 175 p. Tese (Doutoramento em Didática da Matemática), Covilhã, 2019.

PINHEIRO, O. L. *et al.* Teste de Progresso: uma Ferramenta Avaliativa para a Gestão Acadêmica. **Revista Brasileira de Educação Médica**, [s. l.], v. 39, n. 1, p. 68-78, 2015.

PITON-GONÇALVES, J.; ALMEIDA, A. M. Análise da dificuldade e da discriminação de itens de Matemática do ENEM. **Revista Eletrônica de Matemática**, Bento Gonçalves, v. 4, n. 2, p. 38-53, 2018.

PRASHANTI, E.; RAMNARAYAN, K. Ten maxims of formative assessment. **Advances in Physiology Education**, [S.l.], v. 43, p. 99-102, 2019.

REBERTI *et al.* Teste de Progresso na Escola Médica: uma Revisão Sistemática acerca da Literatura. **Revista Brasileira de Educação Médica**, Brasília, v. 44, n. 1, p. 1-9, 2020.

ROSA, M. I. *et al.* O Teste de Progresso como Indicador para Melhorias em Curso de Graduação em Medicina. **Revista Brasileira de Educação Médica**, Brasília, v. 41, n. 1, p. 58-68, 2017.

ROSSI, G. Z. *et al.* Abordagens de aprendizado e sua correlação com ambiente educacional e características individuais em escola médica. **Revista Brasileira de Educação Médica**, Brasília, v. 45, n. 3, p. 1-11, 2021.

SAKAI, M. H. *et al.* Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina. **Revista Brasileira de Educação Médica**, Brasília, v. 32, n. 2, p. 254-263, 2008.

SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item. **Psicologia: Reflexão e Crítica**, São Paulo, v. 26, n. 2, p. 241-250, 2013.

SCHUTTPELZ-BRAUNS, K. *et al.* Comparison of the evaluation of formative assessment at two medical faculties with different conditions of undergraduate training, assessment and feedback. **GMS Journal for Medical Education**, [s. l.], v. 37, n. 4, 2020.

SCOTT, H. K.; COGBURN, M. **Piaget**. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, Jan. 2021.

SCULLY, D. Constructing Multiple-Choice Items to Measure Higher-Order Thinking. **Practical Assessment, Research, and Evaluation**, [s. l.], v. 17, n. 4, p. 1-12, 2017.

SHAH, S. S. H. *et al.* Psychometric Analysis of MCQs Used in Assessing the Students at Entrance to a Medical College. **Annals of King Edward Medical University**, [s. l.], v. 18, n. 3, 2012.

SOOBARD, R.; RANNIKMÄE, M. Examining Curriculum Related Progress Using a Context-Based Test Instrument – A Comparison of Estonian Grade 10 and 11 Students. **Science Education International**, [s. l.], v. 26, n. 3, p. 263-283, 2015.

SPRECHER, E. A. Back to the chalkboard: Lessons in scaffolding using SOLO taxonomy from school teachers for university educators. **Psychology Teaching Review**, [s. l.], v. 25, n. 2, p. 95-102, 2019.

THAMPY, H; WILLERT, E; RAMANI, S. Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid, **Journal of General Internal Medicine**, Philadelphia (PA), v. 34, n. 8, p. 1631-1636, 2019.

VAN DER VLEUTEN, C. P. M.; SCHUWIRTH, L. W. T. Assessment in the context of problem-based learning. **Advances in Health Sciences Education**, [s. l.], v. 24, p. 903-9016, 2019.

VANDERBILT, A. A.; FELDMAN, M.; WOOD, I. K. Assessment in undergraduate medical education: a review of course exams. **Medical Education Online**, Philadelphia (PA), v. 18, n. 1, p. 1-5, 2013.

VEGI, V. A. K. *et al.* Multiple-choice questions in assessment: Perceptions of medical students from low-resource setting. **Journal of Education and Health Promotion**, Philadelphia (PA), v. 11, p. 1-6, 2022.

VIANNA, H. M. **Testes em Educação**, São Paulo: Ibrasa, 6. ed, 1987. 222 p.

VILARINHO, A. P. L. **Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP**. 2015. 99 p. Dissertação (Mestrado em Matemática em Rede Nacional) - Universidade de Brasília, Brasília, 2015.

VILLELA, E. F. M. *et al.* Análise da adequação dos itens do Teste de Progresso em medicina. **Revista Brasileira de Educação Médica**, Brasília, v. 46, p. 1-7, 2022.

VUMA, S.; SA, B. A comparison of clinical-scenario (case cluster) versus stand-alone multiple choice questions in a problem-based learning environment in undergraduate medicine. **Journal of Taibah University Medical Sciences**, Medina, v. 12, n. 1, p. 14-26, 2017.

WEAY, A. L.; MASOOD M.; ABDULLAH, S. H. Systematic Review of Revised Bloom Taxonomy, SOLO Taxonomy and Webb's Depth of Knowledge (DOK) in Assessing Students' Historical Understanding in Learning History. **Malaysian Journal of Higher Order Thinking Skills in Education**, [s. l.], p. 1-26, 2016.

YURTYAPAN, M. I.; YILMAZ, G. K. An Investigation of the Geometric Thinking Levels of Middle School Mathematics Preservice Teachers According to SOLO Taxonomy: "Social Distance Problems". **Participatory Educational Research**, Bursa, v. 8, n. 3, p. 188-209, 2021.

APÊNDICE A - Solicitação de dispensa do Termo de Consentimento Livre e Esclarecido (TCLE)

Declaro, por meio deste documento, que o presente projeto de pesquisa, “AVALIAÇÃO DO TESTE DO PROGRESSO DE UMA FACULDADE DE MEDICINA PELOS PRESSUPOSTOS DA TAXONOMIA SOLO” tem por objetivo avaliar o Teste do Progresso do Curso de Medicina da UNIFENAS-BH no ano de 2022-1 pelos pressupostos da Taxonomia SOLO.

Em virtude de tratar-se de dado secundário (prova/documento), armazenado pela Instituição de Ensino responsável pela aplicação do Teste Progresso, sem que haja qualquer acesso às respostas dos alunos que realizaram os testes ou aos professores que o elaboraram, solicito a dispensa do TCLE, para os fins dessa pesquisa. Além do mais, em virtude do tempo de aplicação, pretérito, seria inviável coletar TCLE dos alunos e professores.

Baseado na importância científica dos dados existentes neste (s) teste (s) justifica-se a análise dos mesmos após a autorização do CEP-UNIFENAS e dos responsáveis pelo arquivamento, com devida assinatura do termo de compromisso para utilização dos dados (TCUD) por parte dos pesquisadores e da instituição. Limitando-se ainda utilizar apenas os dados necessários para o desenvolvimento do presente projeto, sendo esta, uma responsabilidade assumida pelo pesquisador responsável.

Belo Horizonte, 09 de janeiro de 2022



Pedro Paulo Trindade Resende

APÊNDICE B - Termo de Compromisso de Utilização de Dados (TCUD)

IDENTIFICAÇÃO DOS MEMBROS DE PESQUISA:

Nome completo (sem abreviação)	CPF
PEDRO PAULO TRINDADE RESENDE	488.424.786-87
ALEXANDRE ARAÚJO PEREIRA	981.509.360-68
JOSÉ MARIA PEIXOTO NETO	026.710.063-90

Identificação da pesquisa

TÍTULO DA PESQUISA: AVALIAÇÃO DO TESTE DO PROGRESSO DE UMA FACULDADE DE MEDICINA PELOS PRESSUPOSTOS DA TAXONOMIA SOLO

PESQUISADOR: Pedro Paulo Trindade Resende

PESQUISADORES PARTICIPANTES: Dr. Alexandre Araújo Pereira

Dr. José Maria Peixoto Neto

ENDEREÇO: Rua Geraldo Pires Teixeira, 96 Colinas Del Rei

São João del-Rei, MG - CEP: 36307-635

TELEFONE DE CONTATO: (32) 99107 8646

E-MAIL: pedroresende@ufsj.edu.br

pedro.trindade@aluno.unifenas.br

PATROCINADORES: patrocínio próprio

Descrição dos Dados

Serão analisados os itens das provas do Teste do Progresso aplicado no Curso de Medicina da UNIFENAS - BH do ano de 2022 1 de acordo com os princípios da Taxonomia SOLO.

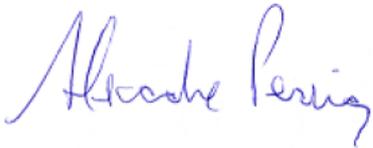
O tempo de pesquisa será de aproximadamente um ano.

Os dados serão coletados somente após aprovação do projeto de pesquisa pelo Comitê de Ética em Pesquisa (CEP) da UNIFENAS.

Declaração dos pesquisadores

Os pesquisadores envolvidos no projeto se comprometem a manter a confidencialidade sobre os dados coletados, bem como a privacidade de seus conteúdos, como preconizam a Resolução nº 466/12, e suas complementares, do Conselho Nacional de Saúde. Declaramos entender que a integridade das informações e a garantia da confidencialidade dos dados e a privacidade dos indivíduos que terão suas informações acessadas estão sob nossa responsabilidade. Também declaramos que não repassaremos os dados coletados ou o banco de dados em sua íntegra, ou parte dele, a pessoas não envolvidas na equipe da pesquisa. Os dados obtidos na pesquisa somente serão utilizados para este projeto. Todo e qualquer outro uso que venha a ser planejado, será objeto de novo projeto de pesquisa, que será submetido à apreciação do CEP Unifenas.

Belo Horizonte, 09 de janeiro de 2022.

Nome completo (sem abreviação)	CPF	Assinatura
PEDRO PAULO TRINDADE RESENDE	488.424.786-87	
ALEXANDRE ARAÚJO PEREIRA	981.509.360-68	
JOSE MARIA PEIXOTO NETO	026.710.063-90	 Assinado de forma digital por JOSE MARIA PEIXOTO:49104292634 Dados: 2022.02.11 08:55:56 -03'00'

APÊNDICE C - Ficha de categorização do item (FCI)

A seguir você encontrará a transcrição da questão de nº do Teste de Progresso da UNIFENAS-BH do ano de 20..... Leia o enunciado da questão e avalie a resposta assinalada como correta. Partindo da resposta correta, analise os conteúdos e os procedimentos cognitivos exigidos para a resolução da questão, considerando o Sistema de Categorização dos Itens apresentado a você. Ao final classifique a complexidade da questão pela Taxonomia SOLO.

Enunciado do item:**Alternativas:****Gabarito da questão:****Análise de conteúdo:****Análise de procedimentos:****Categoria SOLO da questão:** () SOLO 2 () SOLO 3 () SOLO 4 () SOLO 5

APÊNDICE D - Sistema de categorização dos itens pela Taxonomia SOLO

SISTEMA DE CATEGORIZAÇÃO DOS ITENS DO TESTE DO PROGRESSO DA UNIFENAS-BH PELA TAXONOMIA SOLO

Categoria SOLO da questão	Exemplos de verbos de comando e sua relação às categorias SOLO	Conhecimento abordado		Procedimentos cognitivos envolvidos
		Número de tópicos utilizados	Relação dos tópicos	
Abstrato (SOLO 5)	Discutir, estruturar, criar hipóteses, avaliar, raciocinar, teorizar, estimar, criticar, interpretar, prever, refletir, programar, julgar, generalizar, implementar, teorizar	Dois ou mais tópicos	Necessita associação entres os tópicos	indução e/ou dedução; exige identificação de informações relevantes, elaboração de hipóteses e generalizações não comumente discutidas no curso médico
Relacional (SOLO 4)	Explicar, integrar, referir, analisar, comparar, interpretar, desenhar, construir, planejar, resumir, colocar em prática, relacionar, argumentar			indução e/ou dedução; exige identificação de informações relevantes frequentemente discutidas no curso médico
Multiestrutural (SOLO 3)	Descrever, fazer algoritmos, executar, resolver, aplicar, formular, listar, combinar, completar, provar, classificar, enumerar, usar métodos		Utilizados isoladamente	indução e/ou dedução; exige identificação de informações relevantes frequentemente discutidas no curso médico
Uniestrutural (SOLO 2)	Identificar, decidir, contar, calcular, organizar, reproduzir, eleger, encontrar, reconhecer, definir, contar, buscar, escolher, parafrasear	Um único tópico	Não se aplica	indução e/ou dedução; exige identificação de informação relevante frequentemente discutida no curso médico

Fonte: Adaptado de Ceia (2002) e Pereira (2019).

APÊNDICE E - Tabelas SOLO

CLÍNICA		GINECOLOGIA		CIRURGIA		PEDIATRIA		SAÚDE COLETIVA						
ITEM	SOLO	ITEM	SOLO	ITEM	SOLO	ITEM	SOLO	ITEM	SOLO					
1	Q 91	4	25	Q56	4	73	Q15	4	97	Q36	4	49	Q105	2
2	Q95	4	26	Q71	2	74	Q23	4	98	Q27	2	50	Q120	4
3	Q76	4	27	Q57	3	75	Q7	4	99	Q32	2	51	Q101	2
4	Q85	2	28	Q66	2	76	Q14	4	100	Q41	4	52	Q107	3
5	Q81	4	29	Q53	4	77	Q1	4	101	Q40	4	53	Q119	2
6	Q80	4	30	Q49	3	78	Q17	4	102	Q26	4	54	Q115	2
7	Q78	4	31	Q52	3	79	Q19	4	103	Q34	4	55	Q99	2
8	Q89	4	32	Q67	2	80	Q20	4	104	Q47	2	56	Q100	2
9	Q75	4	33	Q50	4	81	Q3	4	105	Q44	4	57	Q98	2
10	Q73	2	34	Q54	4	82	Q11	3	106	Q30	4	58	Q109	2
11	Q88	4	35	Q64	4	83	Q10	3	107	Q33	3	59	Q108	2
12	Q96	4	36	Q68	4	84	Q2	2	108	Q31	4	60	Q116	2
13	Q77	4	37	Q69	2	85	Q16	3	109	Q25	2	61	Q118	2
14	Q74	4	38	Q51	4	86	Q18	4	110	Q39	2	62	Q106	2
15	Q84	4	39	Q60	4	87	Q24	4	111	Q35	2	63	Q97	2
16	Q93	4	40	Q63	4	88	Q9	4	112	Q37	4	64	Q102	3
17	Q92	4	41	Q62	4	89	Q13	2	113	Q43	4	65	Q106	4
18	Q86	4	42	Q55	2	90	Q4	4	114	Q28	4	66	Q114	2
19	Q82	3	43	Q65	2	91	Q91	3	115	Q48	4	67	Q112	4
20	Q87	4	44	Q70	4	92	Q21	4	116	Q38	4	68	Q103	2
21	Q90	4	45	Q59	2	93	Q12	4	117	Q45	4	69	Q104	4
22	Q79	4	46	Q58	4	94	Q5	2	118	Q42	4	70	Q113	2
23	Q94	4	47	Q72	4	95	Q22	4	119	Q29	3	71	Q103	2
24	Q83	4	48	Q61	4	96	Q8	4	120	Q46	2	72	Q117	2

ANEXO A - Link de acesso aos itens do TPUBH2022-1

https://docs.google.com/document/d/1V3SM_cLmPYgfWnprkZfRw16QK0C-WjxR/edit?usp=share_link&oid=103188252538609047663&rtpof=true&sd=true