

## SELEÇÃO DE INSTRUMENTO DE AVALIAÇÃO DO PENSAMENTO CRÍTICO E CRIATIVO: UMA REVISÃO SISTEMÁTICA DA LITERATURA

### *SELECTION OF CRITICAL AND CREATIVE THINKING ASSESSMENT INSTRUMENT: SYSTEMATIC REVIEW OF THE LITERATURE*

**Rosa Sandra Paiva**

ORCID 0000-0002-0155-916X

Centro de Investigação em Didática e Tecnologia na  
Formação de Formadores (CIDTFF)  
Universidade de Aveiro  
Aveiro, Portugal  
sandra.paiva@ua.pt

**António Pedro Costa**

ORCID 0000-0002-4644-5879

Centro de Investigação em Didática e Tecnologia na  
Formação de Formadores (CIDTFF)  
Universidade de Aveiro  
Aveiro, Portugal  
apcosta@ua.pt

**Luís Paulo Reis**

ORCID 0000-0002-4709-1718

Laboratório de Inteligência Artificial e Ciência de  
Computadores (LIACC)  
Universidade do Porto  
Porto, Portugal  
lpreis@fe.up.pt

**Margarita Pino-Juste**

ORCID 0000-0002-4709-1718

Research Group on Education  
Physical Activity and Health (GIES10)  
Universidade de Vigo  
Vigo, Espanha  
mpino@uvigo.es

**Resumo.** A relevância do pensamento crítico e criativo (PCC) na educação é indiscutível, especialmente considerando que avaliações globais indicam que muitos estudantes ingressam no ensino superior com capacidades de pensamento crítico ainda em desenvolvimento e alcançam apenas níveis moderados de proficiência no final dos cursos. Apesar desta importância, o PCC permanece subvalorizado nas práticas educativas e, particularmente, nas avaliações, comprometendo a realização pessoal e profissional dos estudantes, bem como a cidadania ativa, a inovação, o desenvolvimento económico e a inclusão social. Este cenário ressalta a necessidade de desenvolver e adotar instrumentos eficazes e credíveis para avaliar as capacidades de PCC. Este artigo realizou uma revisão sistemática da literatura sobre instrumentos de avaliação do PCC, com o objetivo de identificar na literatura um modelo e um instrumento de avaliação que integrasse as capacidades de PCC e fosse aplicável a alunos do secundário. A pesquisa baseou-se em estudos de 2018 a 2022 e utilizou bases de dados SCOPUS e Web of Science. Os instrumentos foram analisados em termos de confiabilidade, fiabilidade, adequação cultural, abrangência, relevância e níveis educativos. Além disso, foram explorados cenários quotidianos e os tipos de questões. Os resultados revelaram que a maioria dos instrumentos de avaliação são voltados para adultos ou estudantes universitários, sendo poucos adaptados a jovens do ensino secundário. De entre os instrumentos analisados, os de Lopes, Silva e Morais e Franco, Costa e Almeida destacaram-se. Os primeiros apresentam cenários práticos, associam PCC e foram concebidos especificamente para o contexto português. Já os segundos instrumentos abrangem todas as dimensões do Pensamento Crítico (capacidades, disposições e hábitos mentais). A investigação conclui que o instrumento de Lopes, Silva e Morais é o mais adequado para estudos qualitativos no ensino secundário, devido à sua abrangência teórica, originalidade e integração de cenários quotidianos.

**Palavras-chave:** capacidades de pensamento crítico e criativo; disposições de pensamento crítico; hábitos mentais de pensamento crítico; resolução de problemas; interdisciplinariedade

**Abstract.** The relevance of critical and creative thinking (CCT) in education is indisputable, especially given global assessments indicating that many students enter higher education with critical thinking skills still under development and achieve only moderate levels of proficiency by the end of their courses. Despite its importance, CCT remains undervalued in educational practices, particularly in assessments, undermining students' personal and professional development, as well as active citizenship, innovation, economic growth, and social inclusion. This scenario highlights the need to develop and adopt effective, validated instruments for assessing CCT skills. This article presents a



systematic review literature of CCT assessment instruments, focusing on identifying tools suitable for qualitative studies in secondary education. These skills are considered essential for the 21st century, being crucial for solving interdisciplinary problems. The research was based on inclusion and exclusion criteria, covering studies from 2018 to 2022, and utilized databases such as SCOPUS and Web of Science. The instruments were analyzed in terms of reliability, validity, cultural adaptation, scope, relevance, and educational levels. Additionally, real-life scenarios and question types were explored. The results revealed that most assessment instruments target adults or university students, with few adapted for secondary school youth. Among the analyzed instruments, those by Lopes, Silva, and Moraes, as well as Franco, Costa, and Almeida, stood out. The former integrates practical scenarios, combines CCT, and was specifically designed for the Portuguese context. The latter encompasses all dimensions of critical thinking (skills, dispositions, and mental habits). The study concludes that the instrument by Lopes, Silva, and Moraes is the most suitable for qualitative studies in secondary education due to its theoretical comprehensiveness, originality, and integration of real-life scenarios.

**Keywords:** Critical and Creative Thinking Skills; Critical Thinking Dispositions; Critical Thinking Mental Habits; Problem-Solving; Interdisciplinarity

## 1. INTRODUÇÃO

Este artigo tem por finalidade investigar formas de operacionalização do Pensamento Crítico e Criativo (PCC) no âmbito das práticas letivas, com especial enfoque na identificação de modelos de aprendizagem e instrumentos de avaliação que permitam a sua implementação efetiva. A relevância do PCC é amplamente reconhecida na literatura, sendo considerado um pilar fundamental para a realização pessoal e profissional dos estudantes, o exercício da cidadania ativa, a promoção da inovação, o desenvolvimento económico e a inclusão social.

Apesar da sua importância reconhecida em diversas esferas educacionais e sociais, o PCC permanece frequentemente subvalorizado nas práticas pedagógicas, sobretudo nos processos avaliativos. Esta lacuna compromete não apenas os objetivos formativos do sistema educativo, mas também o desenvolvimento de competências essenciais para enfrentar os desafios do século XXI.

Esse cenário evidenciou a necessidade de se efetuar uma Revisão Sistemática da Literatura (RSL), com o objetivo de identificar, na literatura, um modelo e um instrumento de avaliação que integrassem as capacidades de PCC e fossem aplicáveis a alunos do ensino secundário.

A presente investigação foi orientada pela seguinte questão de pesquisa: Qual o modelo e o instrumento de avaliação que integrem as capacidades de PCC, aplicáveis a alunos do ensino secundário? Para responder a esta questão, foi seguido um protocolo metodológico estruturado em diversas fases: definição da pergunta de revisão; elaboração da introdução com a problematização do fenómeno de interesse; definição dos critérios de inclusão e exclusão; estratégias de busca; avaliação da qualidade metodológica dos estudos selecionados; extração e análise dos dados; síntese dos estudos; e meta-agregação dos resultados.

Este artigo encontra-se organizado da seguinte forma: inicialmente, apresenta-se o enquadramento teórico sobre o PCC, incluindo as suas definições, dimensões e relevância educacional. No âmbito da avaliação do PCC, discorrem-se os principais desafios, os instrumentos existentes e a respetiva categorização. Em seguida, descreve-se detalhadamente a metodologia da RSL adotada; posteriormente, são analisados os resultados obtidos; e, por fim, apresentam-se as conclusões e as implicações decorrentes do estudo.

## 2. ENQUADRAMENTO TEÓRICO

### 2.1. Definições de Pensamento Crítico e Criativo

O pensamento crítico (PC) tem sido objeto de intensa exploração teórica, resultando numa multiplicidade de definições ao longo das décadas. O PC é, frequentemente, entendido como

sinónimo de pensamento reflexivo (Zapalska, 2018). De entre as definições mais influentes encontra-se a de Ennis (2011) que descreve o PC como “um pensamento racional e reflexivo focado em decidir no que acreditar ou fazer.” [nossa tradução] (p.1). Já Facione (1990) define o PC como:

The Delphi panel articulated its understanding of CT as follows: We understand CT to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based (p. 4).

Outros teóricos ampliaram esta definição. Halpern (1997) considera o PC como “o uso de competências cognitivas ou estratégias que aumentam a probabilidade de alcançar resultados desejados” [nossa tradução], (p. 4), dando ênfase à maneira de atingir um objetivo, através da divisão do processo do pensamento em partes. Tal divisão é algo artificial, pois não é fácil separar os problemas na vida real, sendo necessário, muitas vezes, estimar probabilidades ao tomar uma decisão ou gerar possíveis soluções numa tarefa de raciocínio. Paul e Elder (2005) afirmam que o pensamento crítico é “*el pensamiento crítico es el proceso de analizar y evaluar el pensamiento con el propósito de mejorarlo*” (p. 7). Saiz & Rivas (2008), por sua vez, descrevem o PC como um “processo de procura do conhecimento através de competências de raciocínio, de resolução de problemas e de tomada de decisões que nos permitam obter os resultados esperados” [nossa tradução] (p. 3), destacando que “raciocinar bem consiste em argumentar com solidez, resolver bem um problema implica alcançar a aplicação da melhor estratégia, e decidir bem exige elaborar juízos precisos de probabilidade e utilizar heurísticas adequadas” [nossa tradução] (p.4).

Esta diversidade de definições de PC e a falta de consenso atraiu a atenção da *American Philosophical Association* (APA), esta identificou as seguintes habilidades como sendo cruciais para esse tipo de pensamento: interpretação, análise, avaliação, inferência, explicação e autorregulação (Nieto & Valenzuela, 2012).

Neste artigo, privilegia-se a definição de Facione (1990).

## 2.2. Dimensões do Pensamento Crítico

O PC pode-se estruturar em três dimensões: competências e habilidades/capacidades; disposições; hábitos mentais. A designação processos de PC inclui as competências ou habilidades (Loyens et al., 2023). No presente estudo usamos a designação de capacidades ao invés de processos. Tais capacidades estão mais ligadas a aspetos cognitivos do PC, enquanto as disposições enquadram-se nos aspetos mais afetivos (Tenreiro-Vieira & Vieira, 2022).

Tendo sido, anteriormente, identificadas as capacidades de PC, pela APA, passamos a deter-nos sobre as disposições. As definições das disposições de PC são várias, sendo frequentemente descritas, por um lado, como propensão, suscetibilidade ou tendência e, por outro lado, como atitudes ou hábitos mentais. Certo é que a maioria dos teóricos do PC apesar de admitirem a complexidade de definirem as disposições, nos seus programas ou modelos, normalmente optam por enfatizar uma das duas conceituações possíveis: ou a atitudinal ou a motivacional (Nieto & Valenzuela, 2012). As disposições são usadas para se referir aos atributos caracterológicos dos indivíduos, constituindo a motivação interna de uma pessoa para agir ou responder a pessoas, eventos ou circunstâncias em condições habituais (Facione, 2000).

A dimensão de hábitos mentais do PC, refere-se às reflexões contínuas sobre as decisões e os problemas quotidianos, revelando uma tendência positiva, ambivalente ou negativa no uso do PC (Facione et al., 2016).

### 2.3. Relevância do Pensamento Crítico

Num universo de pouco consenso sobre o que os alunos devem aprender destaca-se o consenso sobre as capacidades de PC como sendo de vital importância para que os alunos se envolvam de maneira significativa e bem-sucedida com sua comunidade (Davies & Mansour, 2022), atingindo maior qualidade de vida (Hakim & Talib, 2018). No contexto profissional, o desenvolvimento do PC constitui a melhor preparação para favorecer a tomada de decisões, a resolução de problemas e a reflexão sobre o próprio desempenho – necessidades do mercado de trabalho, este é reconhecido pelos estudantes, os governos e os empregadores (Danczak et al., 2020). O uso do PC para uma cidadania informada é uma condição necessária para o sucesso quer das instituições democráticas, quer para as empresas económicas competitivas do mercado livre. Assim, é do interesse da educação que todos os cidadãos possam aprender a pensar criticamente, não apenas para o seu bem pessoal, mas também para o bem de todos nós (Facione, 2013).

O PC é reconhecido como uma das quatro competências fundamentais do século XXI – ao lado do pensamento criativo, colaborativo e comunicativo, designadas por os 4 C's (Helaluddin et al., 2023), o qual requer um contexto de aprendizagem ativa ou centrada no aluno, exigindo uma grande mudança na aprendizagem. Esta aprendizagem deve ser capaz de estimular os alunos a pensar criticamente sobre um material ou problema, a colaborar com os colegas no estudo do material, a ter capacidade criativa na resolução de um problema e a comunicar ideias ou conceitos obtidos, tanto oralmente como através da escrita (Thamrin et al., 2023). A resolução de problemas, por meio de metodologias como a Aprendizagem Baseada em Problemas (PBL), é especialmente eficaz para promover capacidades de pensamento de ordem superior, incluindo a análise, o PC, a resolução de problemas e a criatividade (Gasmi & Al Nadabi, 2023; Nurrijal et al., 2023). “O PC é a forma como se abordam os problemas e as questões, é a melhor forma que conhecemos para chegar à verdade” [nossa tradução] (Facione, 2013, p. 10). A literatura refere a forte ligação entre o PC e o pensamento criativo, para uns autores o PC incentiva e melhora a criatividade (Bradley & Price, 2016), e, para outros, ambos os pensamentos são inseparáveis: por um lado a criatividade domina o processo de fazer ou produzir, por outro lado a criticidade domina o processo de avaliar ou julgar. Assim, ao pensar produz-se e avalia-se, gera-se e julga-se, em simultâneo, os produtos que se constroem (Paul & Elder, 2005). Tanto a literatura de investigação como o discurso político identificaram duas capacidades essenciais: a criatividade e o PC, não só para a realização pessoal e profissional, mas também para a cidadania ativa, a inovação, o desenvolvimento económico e a inclusão social (Bouckaert, 2023).

O pensamento criativo, é definido como uma atividade mental complexa e proposital, orientada pela motivação para procurar causas e soluções para um problema, situação ou fenómeno e alcançar soluções diferentes e originais (Al-Massarweh et al., 2024).

Dada a relevância do PC e do Criativo, e a interdependência entre eles, no nosso estudo centra-se em ambos os pensamentos.

### 2.4. Desafios de Avaliação do Pensamento Crítico

A importância atribuída ao ensino-aprendizagem do pensamento crítico e criativo (PCC) e à sua avaliação ainda é desigual, impactando negativamente no sucesso educativo, pois os docentes e os alunos canalizam os seus esforços para o conteúdo da avaliação predominante. Apesar dos avanços na transição de aulas tradicionais para abordagens centradas no aluno e na inclusão de conteúdos baseados em casos e problemas nos currículos, a inovação esperada nos instrumentos de avaliação não acompanhou essas mudanças. Deste modo, os testes de múltipla escolha permanecem amplamente utilizados, privilegiando a memorização, enquanto as

avaliações abertas, mais alinhadas aos objetivos reformistas, ainda são subutilizadas (Van Damme et al., 2023).

No âmbito das orientações políticas e dos padrões de qualificação que regem o Ensino Superior (ES), a criatividade e o PC são amplamente enfatizados. Contudo, essas capacidades são escassamente integradas nas dimensões das avaliações centralizadas administradas no nível dos sistemas (Bouckaert, 2023). Paralelamente, o crescente reconhecimento da relevância do PC impulsionou o interesse por testes padronizados em instituições acadêmicas. Esses instrumentos oferecem uma alternativa para avaliar as capacidades que as avaliações tradicionais frequentemente falham. Além disso, destacam-se por serem um meio relativamente barato, fácil de administrar e, pelo menos superficialmente, terem o objetivo de avaliar competências essenciais (Rear, 2018).

Embora os instrumentos padronizados tenham limitações em medir competências avançadas, como a criatividade e o PC, devido ao seu formato, eles incentivam as instituições, os professores e os alunos a se concentrarem em alcançar bons resultados nessas avaliações. Olhando para o futuro, tais instrumentos podem servir como modelos inspiradores para práticas de avaliação mais desenvolvidas e personalizadas por professores ou instituições (Bouckaert, 2023).

A validação de instrumentos de avaliação de PC tem aberto um debate considerável. Atualmente, os testes padronizados que se baseiam em taxonomias de capacidades de PC bem estabelecidas, utilizados há décadas, têm sido questionados sobre a transferibilidade da validade do constructo entre disciplinas ou regiões (Rear, 2018).

## 2.5. Instrumentos de avaliação do Pensamento crítico

A avaliação do PC conta com diversas ferramentas teóricas e instrumentais. Entre as mais utilizadas estão as taxonomias educacionais, como a Taxonomia de *Bloom* revista e a Taxonomia SOLO (*Structure of the Observed Learning Outcome*), além de testes padronizados amplamente reconhecidos, como o *California Critical Thinking Skills Test* (CCTST), o *California Critical Thinking Disposition Inventory* (CCTDI) e o *Watson-Glaser II Critical Thinking Appraisal* (WGCTA). Há uma variedade de instrumentos de avaliação, uns centram-se apenas nas capacidades, outros nas disposições e outros em ambas. Contudo, é relevante que os estudos se foquem em ambas, uma vez que o momento de desenvolvimento deve ser em simultâneo, pelo que estas são mutuamente reforçadas, embora outros defendam que é necessário cultivar as disposições antes das capacidades (Facione, 1995).

### 2.5.1. Taxonomia de *Bloom* revista

A taxonomia de *Bloom* revista é uma estrutura bidimensional, por ter uma dimensão de conhecimento e uma dimensão dos processos. A primeira dimensão apresenta quatro categorias: Efetivo/Factual, Conceptual, Procedural e Metacognitivo. A segunda dimensão reúne outras seis categorias: Lembrar, Entender, Aplicar, Analisar, Sintetizar e Criar. Esta taxonomia mantém o design hierárquico da original, mas é flexível, pois considera-se a possibilidade de interpolação das categorias do processo cognitivo quando necessário (Ferraz & Belhot, 2010). Os três níveis mais baixos são: conhecimento, compreensão e aplicação, e os três níveis superiores são: análise, síntese e avaliação, sendo que os referidos níveis de ordem superior representam o PC (Faravani & Taleb, 2020). A taxonomia de *Bloom*, a versão de 1971 e a versão revista de 1974, privilegiam o desenvolvimento passo a passo das capacidades do PC, a resolução de problemas e de questões reais, a indicação de instruções claras e inequívocas, a estruturação de relatórios e projetos (Zapalska et al., 2018).

### 2.5.2. Taxonomia SOLO





*A Structure of the Observed Learning Outcome* (taxonomia SOLO) desenvolvida por Biggs e Collis, usada para medir a capacidade de pensamento/raciocínio, nomeadamente na área de raciocínio de alto nível: pensamento criativo, pensamento lógico e PC (Damopolii et al., 2019). A taxonomia SOLO foi projetada para classificar as diferenças entre conhecimento superficial e profundo. É uma estrutura hierárquica com cinco níveis, os alunos devem articular os seus conhecimentos usando o vocabulário de cada nível, para garantir a compreensão e a aprendizagem significativa (Davies & Mansour, 2022).

### 2.5.3. Testes padronizados de Pensamento Crítico

Para a avaliação do PC foram construídos vários instrumentos, contudo a sua confiabilidade e a sua validade estão a ser debatidas por instituições como a APA, pois apresentam alguns problemas conceituais e metodológicos relativos ao conceito de PC (Ossa-Cornejo et al., 2017).

Entre os testes padronizados de PC os mais comumente usados são: o *California Critical Thinking Skills Test* - CCTST, o *California Critical Thinking Disposition Inventory* - CCTDI e o *Watson-Glaser II Critical Thinking Appraisal* - WGC-TA (Ross et al., 2013).

#### California Critical Thinking Skills Test (CCTST)

O CCTST tem por base a conceptualização consensual do PC advinda do projeto de investigação Delphi patrocinado pela APA (Loyens, 2023). Este teste é amplamente utilizado para medir as capacidades de PC, apresentando uma confiabilidade de 0,78 a 0,80 (Ossa-Cornejo et al., 2017). O instrumento fornece uma pontuação geral e cinco subescalas que avaliam interpretação, análise, avaliação, explicação, inferência e autorregulação (Ross et al., 2013). O CCTST é um preditor significativo do sucesso académico, para estudantes do primeiro ao terceiro ano, especialmente em estágios e cursos práticos (Choi et al., 2018).

#### California Critical Thinking Disposition Inventory (CCTDI)

O CCTDI mede as disposições do PC, como mente aberta, analiticidade, curiosidade, busca pela verdade e maturidade (Yue et al., 2017). Desenvolvido, também, sob o projeto Delphi, o teste busca avaliar a motivação dos indivíduos para utilizar capacidades críticas (Facione, 1990). Regista uma confiabilidade de 0,72 a 0,80, com alfa de Cronbach de 0,90 (Ossa-Cornejo et al., 2017). Verifica-se uma grande falta de medidas confiáveis e válidas para avaliar as disposições. Atualmente, as únicas medidas explícitas de disposições de PC disponíveis são a CCTDI de Facione e a, mais recentemente desenvolvida, *Critical Thinking Disposition Scale* (CTDS) de Sosu (Quinn et al., 2020).

A maior contribuição deste instrumento reside no renascimento do interesse no lado disposicional do pensamento, vindo combater a ênfase excessiva dada às capacidades de PC, (Facione et al., 2000). CCTDI foi, também, considerado um preditor significativo para estudantes do primeiro ao terceiro ano (Choi et al., 2018).

#### Watson-Glaser II Critical Thinking Appraisal (WGC-TA)

O WGC-TA, na versão original com 80 questões, regista uma confiabilidade de 0,73-0,83. Na versão reduzida, com 40 questões, regista uma confiabilidade de 0,82 (medida com alfa de Cronbach). Este teste apresenta cinco subescalas: inferência; reconhecimento de pressupostos; dedução; interpretação; avaliação (Ossa-Cornejo et al., 2017) e gera uma única pontuação (Ross et al., 2013). Apesar de sua ampla utilização, o WGCTA não previu o sucesso académico dos estudantes (Choi et al., 2018).

## 2.6. Categorização dos instrumentos de avaliação do Pensamento Crítico

Há uma diversidade de instrumentos de avaliação, de uma forma genérica, propomos categorizá-los para análise quanto: à idade ou nível de ensino em que se aplicam, a existência ou ausência de cenário da vida real e os tipos de questões apresentadas – de resposta aberta, fechada ou mista.

### 2.6.1. Idade ou nível de ensino

A grande maioria dos instrumentos de avaliação de PC destinam-se a adultos ou estudantes universitários, havendo poucos testes específicos para estudantes mais jovens, embora os Cornell Critical Thinking Test (CCTT) apresentem testes para crianças e jovens.

O CCTT apresenta algumas inconsistências nas subescalas reportadas em investigações e na literatura. A dificuldade em obter os manuais originais de 1985, bem como a versão revista de 2005, limita a análise detalhada. (Loyens et al., 2023).

O nível X é dirigido a crianças e jovens entre 9 e 18 anos (Ossa-Cornejo et al., 2017) e abrange os processos de Indução, Dedução, Observação, Credibilidade e Suposição, contudo essas cinco dimensões são reduzidas a quatro no manual de teste (Loyens et al., 2023). Por sua vez, o nível Z é destinado a estudantes universitários e adultos, mas também pode ser aplicado a estudantes avançados do ensino básico e secundário. Este nível avalia habilidades como indução, credibilidade de fontes, semântica, predição e experimentação, falácias, dedução, definição e identificação de hipóteses (Ossa-Cornejo et al., 2017).

Apesar de o CCTT, nos seus níveis X, Y e Z, ser parcialmente adaptável a jovens, outras propostas necessitam de maior consolidação psicométrica (Manassero-Mas & Vázquez-Alonso, 2024).

### 2.6.2. Cenário

A apresentação de cenários confere uma dimensão mais concreta ao instrumento, mais próxima da prática, a relevância destes cenários é destacada com a seguinte descrição: “O teste HCTA é considerado um ponto de viragem no cenário de avaliação do PC, principalmente pelo seu O formato, pela apresentação de cenários da vida real” (Franco, Costa & Almeida, 2018, p. 293).

O Halpern Critical Thinking Assessment using (HCTA), apresenta 25 situações-problema, nas quais é feita uma pergunta de resposta fechada e, em seguida, uma de resposta aberta (Ossa-Cornejo et al., 2017).

### 2.6.3. Tipo de questões

Os testes de PC são, habitualmente, divididos em dois grupos, os que usam itens de múltipla escolha, com respostas fechadas; e aqueles que recorrem a respostas abertas ou de redação (Ossa-Cornejo et al., 2017). Muitas avaliações de PC dependem exclusivamente do reconhecimento (por exemplo, múltipla escolha e classificação), enquanto outras dependem exclusivamente da recordação (por exemplo, resposta curta e redação) (Butler et al., 2012). Ambos os grupos de avaliações têm aspetos positivos, por um lado os testes de resposta fechada apresentam bons resultados estatísticos (validade e confiabilidade), por outro lado os testes de resposta aberta são mais adaptáveis às necessidades específicas de cada caso ou definição e são mais fáceis de repetir. Também têm aspetos negativos, os testes de resposta fechada apenas avaliam aspetos predeterminados do PC, apresentam dificuldades para sua repetição, os problemas usados são artificiais e distantes das atividades diárias das pessoas. Os testes de resposta aberta têm uma validade limitada ao nível do construto e da confiabilidade e têm

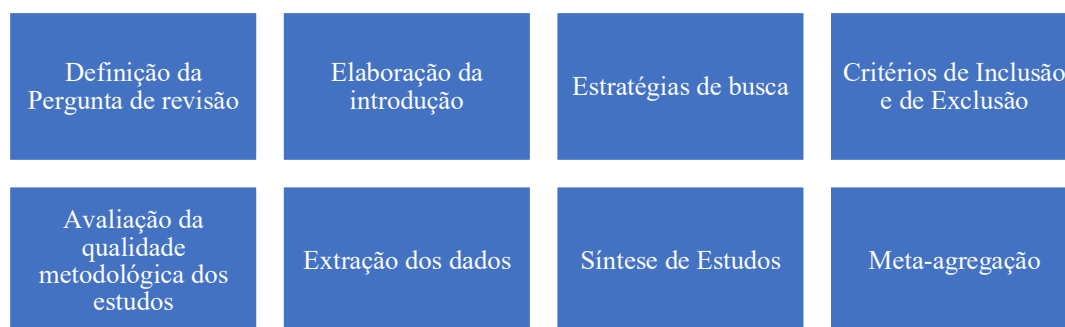
custos mais elevados aliados ao tempo longo de correção, pois exigem avaliadores humanos e especialistas (Ossa-Cornejo et al., 2017).

Também existem alguns testes híbridos, incluem os dois tipos de questões, como é o caso do Teste de Habilidades de Pensamento Crítico Halpern (Ossa-Cornejo et al., 2017).

Enquadram-se no grupo de avaliação por reconhecimento, por exemplo, as seguintes avaliações: as Measures of Academic Proficiency and Progress; o California Critical Thinking Skills Test (CCTST); o Cornell Critical Thinking Test (CCTT). No grupo de avaliação por recordação, por exemplo, incluem-se: o ICAT Critical Thinking Essay Examination, o Cornell (também conhecido como Ennis-Weir) Critical Thinking Essay, e o Watson-Glaser II Critical Thinking Appraisal (WGC-TA) (Butler et al., 2012).

### 3. METODOLOGIA

Para realizar este estudo, utilizou-se como procedimento metodológico uma Revisão Sistemática da Literatura (RSL). Este método “destina-se a conhecer (e dar a conhecer) o estado de arte, sistematizando a informação obtida e transformando-a em conhecimento aprofundado sobre um dado tema” (Cardoso, Alarcão e Celorico, 2010, p. 26), sendo muito úteis “nas áreas sociais e na saúde (...) como geradoras de sínteses de evidências para informar a tomada de decisão no âmbito das práticas e das políticas” (Soares, et al., 2021, p.13). As RSL de alta qualidade são consideradas mais fiáveis, no entanto são demasiado demoradas, particularmente para “aqueles que exigem sínteses de evidência para a tomada de decisões vivem cada vez mais em tempos de ritmo mais rápido, influenciados por intervenções inovadoras e tecnologias que aceleram a comunicação e a interação” (Schünemann & Moja, 2015, p. 1). A fim de garantir a referida qualidade, para presente investigação definiram-se várias fases que integram o protocolo de RSL. Estas são: Definição da Pergunta de revisão; Elaboração da introdução – para problematizar o fenómeno de interesse; Critérios de Inclusão e de Exclusão; Estratégias de busca; Avaliação da qualidade metodológica dos estudos; Extração dos dados; Síntese de Estudos; Meta-agregação.



**Figura 1.** Fases do Protocolo de Revisão Sistemática de Literatura

Fonte: Elaborado a partir de Soares, et al. (2021)

#### 3.1. Definição da Pergunta de revisão

A definição da pergunta de revisão seguiu o mnemónico PICO definido por Joanna Briggs Institute (JBI) para ajudar a formular a pergunta de revisão (Soares, et al., 2021). O “P” se refere aos participantes e as suas características, estes são estudantes portugueses adolescentes do Ensino Secundário, pertencentes ao curso de Curso de Línguas e Humanidades. O “I” refere-se ao fenómeno de interesse, isto é, a identificação de instrumentos de avaliação do PC. Por fim, o “Co”, identifica elementos importantes do contexto, de natureza educativa centrado na resolução de problemas reais interdisciplinares. Deste processo resultou a seguinte questão de



investigação: Qual o modelo e o instrumento de avaliação que integre as capacidades de PCC aplicável a alunos do secundário?, enquadrada na questão de investigação mais ampla: Em que medida a utilização do ChatGPT como ferramenta de suporte afeta o desenvolvimento das capacidades e disposições de pensamento crítico e criativo em alunos do ensino secundário? Neste estudo interessou identificar o modelo de pensamento crítico e criativo antes de incluir a dimensão da Inteligência Artificial Generativa através do ChatGPT. Assim, o objetivo desta RSL foi o de identificar e de selecionar na literatura um modelo que integrasse as capacidades de Pensamento Crítico e Criativo aplicável a alunos do secundário.

### 3.2. Elaboração da introdução

A introdução foi elaborada com o objetivo de apresentar uma base teórica sólida e abrangente sobre o conceito de PC, as suas dimensões, a relevância, os desafios, os instrumentos de avaliação e tipos de instrumentos de avaliação

Relativamente ao conceito do PC, na introdução contextualiza-se o PC como um constructo teórico multifacetado, explorado por diferentes autores ao longo das décadas. Definições de pensadores como Ennis (2011), Facione (1990) e Halpern (1997) são apresentadas para ilustrar as diferentes perspetivas sobre o tema. Esses conceitos convergem para descrever o PC como um processo cognitivo que envolve análise, avaliação e resolução de problemas. Essa diversidade de definições reflete a complexidade do PC e fundamenta o uso de variadas abordagens teóricas nos instrumentos de avaliação.

Sobre as dimensões do PC, na introdução o PC é apresentado como sendo composto por três dimensões principais: capacidades (skills ou abilities), disposições e hábitos mentais. Essa visão segmentada do PC é essencial para compreender com maior profundidade do PC e sua aplicabilidade prática. As capacidades estão ligadas aos aspetos cognitivos, enquanto as disposições refletem fatores afetivos e motivacionais, e os hábitos mentais abrangem as reflexões contínuas sobre o uso do PC. Esses elementos fornecem uma estrutura teórica para categorizar os instrumentos de avaliação analisados.

Na introdução, a relevância do PC, é destacada em múltiplos contextos: educacional, profissional e social. O PC é reconhecido como uma das competências fundamentais do século XXI, ao lado do pensamento criativo, colaborativo e comunicativo. Também se destaca como o PC é vital para a resolução de problemas e a inovação, reforçando sua relevância para o sucesso pessoal e profissional. O pensamento criativo, por sua vez, complementa o PC, estimulando a geração de ideias e soluções originais, o que torna ambos indispensáveis para enfrentar os desafios contemporâneos. Essa fundamentação justifica a necessidade de instrumentos eficazes que avaliem tanto o PC quanto o pensamento criativo.

A introdução explora os desafios na avaliação do PC, destacando a desfasamento entre os avanços pedagógicos e os instrumentos de avaliação. Enquanto metodologias centradas no aluno, como abordagens baseadas em problemas, têm evoluído, os testes de múltipla escolha continuam a predominar, privilegiando a memorização em detrimento de avaliações abertas mais alinhadas às competências atuais. Este desfasamento estende-se às orientações políticas e dos padrões de qualificação que enfatizam amplamente a criatividade e o PC, mas apresenta uma escassa integração destas capacidades nas avaliações administradas ao nível do sistema. É, também, destacado o uso exponencial dos testes padronizados de PC, apesar da validade e a transferibilidade desses instrumentos continuam em debate.

Relativamente Instrumentos de Avaliação do PC, a introdução aborda as principais taxonomias e instrumentos utilizados para avaliar o PC, como a Taxonomia de Bloom Revista, a Taxonomia SOLO e os testes padronizados amplamente reconhecidos, como o CCTST, CCTDI e WGCTA. É destacada a necessidade de instrumentos que contemplem tanto as capacidades quanto as disposições, conforme defendido por Facione (1995). Essa visão serve como base para a análise dos instrumentos específicos abordados no estudo.

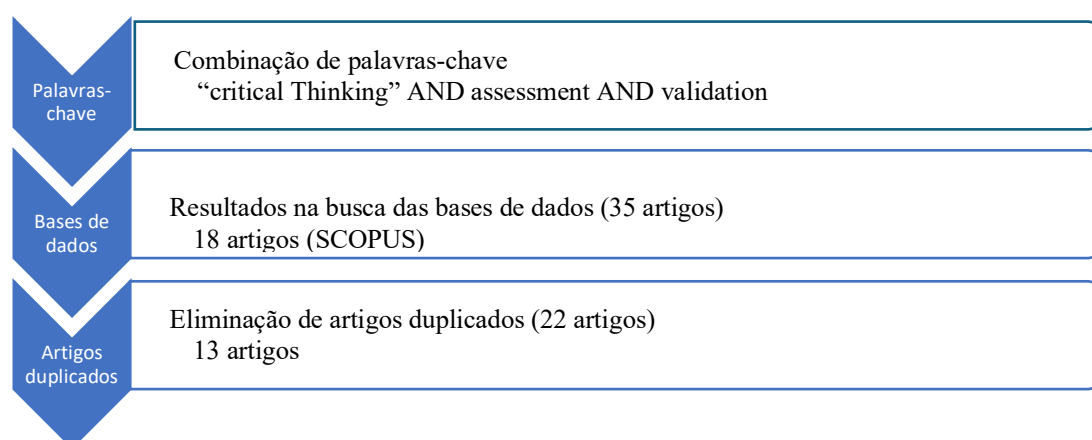
A introdução organiza os instrumentos de avaliação do PC com base em diferentes critérios: idade ou nível de ensino, cenários da vida real e tipos de questões. No que diz respeito à idade ou nível de ensino, destaca-se que a maioria dos instrumentos foi desenvolvida para adultos ou estudantes universitários, com poucos testes específicos para jovens. Adicionalmente, a introdução sublinha a importância da apresentação de cenários da vida real no contexto de avaliação do PC. Quanto aos tipos de instrumentos de avaliação do PC são organizados com base na classificação por tipo de resposta. Os instrumentos são apresentados como testes de resposta fechada, que dependem do reconhecimento, e testes de resposta aberta, que recorrem à recordação através de redações ou respostas curtas. Além disso, são referidos testes híbridos, que combinam características de ambos os formatos, oferecendo maior flexibilidade avaliativa. A introdução, também, discute as vantagens e limitações desses instrumentos.

### 3.3. Estratégias de busca

As estratégias de busca referem-se às bases de dados utilizadas, à delimitação do período e aos termos de busca. As bases elegidas são de natureza multidisciplinar: a SCOPUS e a *Web of Science* e o período delimitado por 5 anos (2018 – 2022), período imediatamente anterior ao surgimento da Inteligência Artificial Generativa (GenAI), razão pela qual esta investigação não se centrou nesta. Não foram incluídos os anos de 2023 e 2024 pela intenção de compreender tendências, impactos e contribuições científicas que precederam a introdução massiva da GenAI. Essa delimitação permite uma análise consistente e evita interferências de fenómenos tecnológicos ainda em fase de consolidação nos anos mais recentes. As palavras-chave usadas foram: “critical Thinking”; assessment; validation, com o boleano AND entre cada palavra, incluíram-se os resultados de acesso aberto e o seguinte tipo de documentos: artigos, artigos de conferência e capítulos de livros. Especificamente SCOPUS, acionamos os filtros relativos às seguintes áreas de estudo: Ciências Sociais, Psicologia e Artes e Humanidades, na Web of Science, elegemos as categorias mais próximas das da SCOPUS: Educação e Pesquisa Educacional; Psicologia Multidisciplinar e Psicologia.

As palavras-chave usadas referem-se ao *Fenômeno de Interesse*, os filtros acionados referem-se ao *Contexto* e relativamente aos *Participantes* não foi definida qualquer estratégia de busca para não restringir mais os resultados.

A pesquisa de artigos, como referido anteriormente, obedeceu à combinação de palavras-chave resultando em 35 artigos, sendo 18 da SCOPUS e 17 da *Web of Science*, contabilizando-se os 13 artigos repetidos obtivemos 22 artigos, como se observa na Figura 2.



**Figura 2.** Número de artigos obtidos de acordo com o Protocolo de RSL – parte 1

Fonte: Adaptada de Franco et al. (2021)

### 3.4. Critérios de Inclusão e de Exclusão

A seleção dos estudos efetuou-se, em primeiro lugar, a partir da leitura do título e do resumo, seguindo os Critérios de Inclusão e de Exclusão. Estes foram definidos em conformidade com os elementos do PICO - População, Intervenção, Contexto e Resultados/Saídas.

#### 3.4.1. Participantes

Relativamente aos *Participantes*, definiram-se os seguintes critérios:

- Instrumentos de Avaliação de PC concebidos para alunos desde crianças a adultos – critério de inclusão;
- Instrumentos de Avaliação usados para avaliar alunos na transição para a vida adulta ou adultos em contexto de trabalho – critério de exclusão.

#### 3.4.2. Fenómeno de Interesse

Os critérios definidos no âmbito do *Fenómeno de Interesse* foram:

- Desenvolvimento de um Instrumento de Avaliação centrado no PC – critério de inclusão;
- Desenvolvimento de um Instrumento de Avaliação centrado noutras capacidades além do PC – critério de exclusão.

#### 3.4.3. Contexto

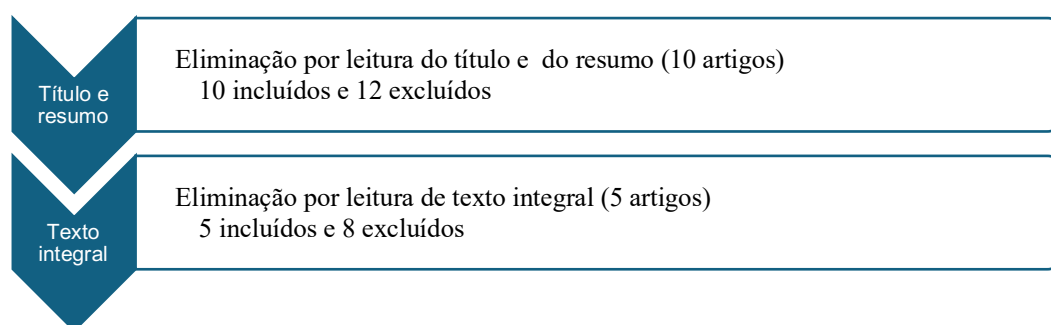
Os critérios referentes ao *Contexto* foram:

- Instrumentos de Avaliação gerais, sem a integração de conteúdos de qualquer disciplina - critério de inclusão;
- Instrumentos de Avaliação específicos, com a integração de conteúdos da área disciplinar STEM ou a disciplinas de ciências ou de desporto – critério de exclusão.

A seleção dos estudos prosseguiu a partir da leitura integral dos estudos e uma pré-análise de conteúdo no webQDA – Qualitative Data Analysis Software (Costa et al., 2019). Desta pré-análise resultou a definição de mais um critério de inclusão e um de exclusão, referentes ao *Fenómeno de Interesse*:

- Desenvolvimento de um Instrumento de Avaliação em que o PC constitui a categoria principal – critério de inclusão;
- Desenvolvimento de um Instrumento de Avaliação em que o PC constitui uma subcategoria – critério de exclusão.

A seleção de artigos obedeceu a critérios de inclusão e de exclusão, mencionados anteriormente. Primeiramente, obtivemos como resultado, a partir da leitura do título e do resumo, 10 artigos, tendo sido 12 os excluídos. Posteriormente, a partir da leitura integral dos artigos obtivemos 5 artigos, sendo 8 excluídos, 7 excluídos atendendo aos últimos critérios definidos relativos ao *Fenómeno de Interesse* e, ainda, 1 excluído considerando, de novo, os critérios de *Contexto*, como se verifica na Figura 3.



**Figura 3.** Número de artigos obtidos de acordo com o Protocolo de RSL – parte 2  
Fonte: Adaptada de Franco et al. (2021)

### 3.5. Avaliação da qualidade metodológica dos estudos

Os critérios de avaliação da qualidade metodológica dos estudos correspondem aos resultados obtidos da validação de cada instrumento. Para os Instrumentos de Avaliação foram escolhidas as dimensões de Confiabilidade e de Fiabilidade. Para os Modelos de Aprendizagem foi escolhida, também, a dimensão da Eficácia. Será aprofundada mais à frente a qualidade metodológica como sendo também uma categoria indutiva – avaliação da qualidade do instrumento.

### 3.6. Extração de dados

A extração de dados implicou selecionar, codificar e inserir informações de cada fonte para armazenamento. Primeiramente, extraíram-se e codificaram-se as informações descritivas, seguindo as categorias dedutivas: autores, ano, idioma do instrumento e o tipo de trabalho - original; traduzido e adaptado para português; traduzido para português; redução de um original.

A pertinência da informação relativa ao tipo de trabalho deve-se ao valor da dimensão cultural de instrumento original concebido para um país com a sua própria cultura ou de um simplesmente traduzido com a presença cultural de outro país. O processo de tradução exige grande precisão e uma adaptação cultural a fim de garantir equivalência linguística e respeito pelas singularidades culturais (Franco et al., 2018). O estudo transcultural engloba a tradução, a adaptação e, ainda, a sua validação (Franco, et al., 2021),

Em segundo lugar, as informações de cada artigo destacaram-se no webQDA (Costa et al., 2019), de forma indutiva, e foram elaboradas categorias de análise, na funcionalidade Códigos Árvore. As categorias foram: a) avaliação da qualidade; b) dimensões do PC; c) bases teóricas; d) níveis de ensino/idades; e) tipos de cenários; f) tipos de questões. Estas categorias constituem as ideias-chave e estruturantes, a partir de cada uma delas foram definidos os respectivos códigos. A categoria a) inclui os códigos: confiabilidade e fiabilidade. A b) engloba os códigos: disposições; capacidades e hábitos mentais. A c) distingue os códigos das bases teóricas de: Facione; Taxonomia de Bloom revista, Taxonomia de Facione (1990b) e Taxonomia de Guilford (1986); Halpern. A categoria d) inclui os códigos: Ensino Superior; Alunos a partir dos 15 anos e Ensino Superior; Alunos a partir dos 17 anos e Ensino Superior. A categoria e) reúne os códigos: sem cenário; 1 cenário uma situação-problema atual e comum do quotidiano; 25 cenários da vida quotidiana de diversas áreas (educação, política, saúde, finanças, ...). Por fim, a categoria f) inclui os códigos: resposta aberta (longa e curta); resposta fechada (múltipla escolha e Escala Likert); resposta mista.

### 3.7. Síntese de Estudos

A síntese de estudos centrou-se em destacar as forças e as limitações de cada instrumento de avaliação do PC, considerando a natureza do próprio instrumento e as características individuais. Este enfoque permitiu identificar os instrumentos mais robustos atendendo aos seus processos de elaboração, tradução ou adaptação – referente à secção do tipo de trabalho e atendendo à confiabilidade ou à fiabilidade – relativa à secção da avaliação da qualidade. Além de ter permitido identificar os instrumentos de melhor aplicabilidade tendo por base os níveis específicos de ensino e os cenários da vida real integrados – relativo à secção de níveis de ensino/idades e os tipos de cenários. Foram, também, avaliadas as abordagens metodológicas e as bases teóricas subjacentes – referente às secções de base teórica e tipos de questões proporcionando uma visão abrangente sobre as potencialidades e restrições de cada instrumento.

A síntese foi segmentada para simplificar a análise, sendo cada parte referente às dimensões do PC- capacidades, disposições e hábitos mentais. Esta segmentação permitiu destacar, com maior ênfase, as forças e as limitações dos instrumentos.

### 3.8. Meta-agregação

A meta-agregação teve como finalidade produzir afirmações que sustentam as tomadas de decisão para a presente investigação. Esse processo baseou-se na análise das características, forças e limitações dos instrumentos avaliados, identificando critérios essenciais para a seleção de ferramentas adequadas à avaliação do pensamento crítico (PC). Os critérios definidos incluíram abrangência, adequação cultural, validação, desenho de cenários, natureza da investigação, generalismo e extensão. A meta-agregação permitiu destacar os instrumentos mais adequados às necessidades específicas da investigação, valorizando características como a integração do pensamento criativo, a aplicabilidade prática e a adaptação cultural ao público-alvo, para além do generalismo dos instrumentos - sem dependência de conteúdos específicos de uma disciplina.

## 4. RESULTADOS

A presente secção aborda as características específicas dos instrumentos de avaliação do pensamento crítico (PC), organizados por diversos critérios, incluindo o tipo de trabalho, a avaliação da qualidade, as dimensões avaliadas, a base teórica, os níveis de ensino, os tipos de cenários, tipos de questões. São, também, apresentadas as forças e as limitações dos instrumentos e os fundamentos para a seleção destes. Esta análise detalhada permite compreender não apenas as especificidades dos instrumentos, mas também as suas aplicações práticas e a sua adequação para diferentes contextos e populações.

### 4.1. Instrumentos de Avaliação de PC – Tipo de Trabalho

Da primeira extração de dados, obteve-se informações descritivas, como referido anteriormente. O idioma de todos os Instrumentos de Avaliação do PC é o português. Apenas os Instrumentos de Avaliação de PC de Lopes et al. (2019; 2021) são originais concebidos para estudantes portugueses. Os Instrumentos de Avaliação de PC de Franco, et al. (2018; 2021) foram traduzidos e adaptados para português, destaca-se que os primeiros autores procederam, também, à validação transcultural. Os Instrumentos de Avaliação de PC traduzidos para português foi, somente, o de Payan-Carreira et al. (2022), o qual também foi reduzido a partir do original. Esta análise encontra-se na Tabela 1.

Estes resultados corroboram a conclusão de que, no contexto português, apenas há um teste de avaliação pensado e construído de raiz para ser utilizado em Portugal que mede o PC – o



teste de PC e Criativo de Lopes et al. (2019). Habitualmente recorre-se a traduções e/ou adaptações de testes internacionais, como é o caso do Cornell *Critical Thinking Test - Level X* e o *Halpern Critical Thinking Assessment* (Lopes et al., 2019). A validação de instrumentos de avaliação de PC tem aberto um debate considerável. Atualmente, os testes padronizados que se baseiam em taxonomias de capacidades de PC bem estabelecidas, e têm sido utilizados há muito tempo, têm sido questionados sobre a transferibilidade da validade do constructo entre disciplinas ou regiões (Rear, 2018). Esses questionamentos sublinham a necessidade de maior atenção à adaptação cultural e à construção de instrumentos localmente relevantes.

**Tabela 1.** Tipo de trabalho desenvolvido nos Instrumentos de Avaliação de PC

Autores	Instrumentos de Avaliação de PC				
	Idioma	Original	Traduzido e adaptado – português	Traduzido - português	Reduzido
Payan-Carreira, et al. (2022)	Português			X	X
Lopes, Silva & Morais (2021)	Português	X			
Franco, et al. (2021)	Português		X *		
Lopes, Silva & Morais (2019)	Português	X			
Franco, Costa & Almeida (2018)	Português		X		

\*Acresce a validação transcultural.

Fonte: dos autores

#### 4.2. Instrumentos de Avaliação de PC – Avaliação da qualidade

A avaliação da qualidade dos Instrumentos de avaliação do PC foi efetuada ao nível da confiabilidade e da fiabilidade, em função das opções de cada artigo. A Confiabilidade para o Instrumento de Avaliação de: Payan-Carreira et al. (2022) foi elevada, com  $\alpha=0,969$ ; Franco, Costa & Almeida (2018) foi aceitável, para o “Pensamento crítico total” (CT total), com  $\Omega=.75$  e para a “recordação livre de PC” (CTfree), com  $\Omega=.70$ . e foi, também, bastante satisfatório para o “reconhecimento de PC” (CTrec), com  $\Omega=.58$ . A Fiabilidade para o Instrumento de Avaliação de: Lopes et al., (2021) foi boa, com uma fiabilidade compósita do construto (FC) de  $FC \geq 0,7$ ; Lopes et al.(2019) altamente significativa, com 0,68 (todas as dimensões  $>0,70$  “bom” a “muito bom”). A Tabela 2 apresenta os valores referidos.

A avaliação da qualidade dos instrumentos, particularmente, aqueles que não são feitos de raiz, isto é, aqueles que são traduzidos e/ou adaptados é analisada dada a sua relevância e a ênfase dada no debate em instituições como a American Colleges and Universities Association. Pelo que a confiabilidade e a validade apresentam alguns problemas conceituais e metodológicos relativos ao conceito de PC (Ossa-Cornejo et al., 2017).

**Tabela 2.** Avaliação da qualidade dos Instrumentos de Avaliação de PC

Autores	Instrumentos de Avaliação do PC	
	Confiabilidade	Fiabilidade
Payan-Carreira, et al. (2022)	Elevada $\alpha=0,969$	

Lopes, Silva & Morais (2021)		Boa FC $\geq 0,7$
Franco, et al. (2021)	-	-
Lopes, Silva & Morais (2019)		Altamente significativa 0,68 Todas as dimensões $>0,70$ (“bom” a “muito bom”)
Franco, Costa & Almeida (2018)	Aceitável CTtotal $\Omega=.75$ ; CTfree $\Omega=.70$ ; CTrec $\Omega=.58$	

Fonte: dos autores

### 4.3. Instrumentos de Avaliação de PC – Dimensões do PC

Os resultados reuniram um conjunto diversificado de instrumentos de avaliação que abrangem as três dimensões do pensamento crítico (PC): capacidades, disposições e hábitos mentais.

O PC pode ser avaliado em diferentes dimensões: capacidades, disposições e hábitos mentais. Os Instrumentos de Avaliação centrados nas capacidades e nas disposições de PC foram os artigos de: Lopes et al. (2021; 2019) e o de Franco et al. (2018). Destaca-se que Lopes et al. (2019) associou capacidades de pensamento criativo às capacidades de PC. Destaca-se, igualmente, que os últimos autores também apresentaram um instrumento centrado nos hábitos mentais (Franco, et al., 2021). A referida associação do Pensamento Criativo ao PC constitui uma mais-valia, dado que ambos estão interligados, o PC incentiva e melhora a criatividade (Bradley & Price, 2016). Ambos os pensamentos, são inseparáveis, por um lado a capacidade crítica domina o processo de avaliar ou julgar, por outro lado a capacidade criativa domina o processo de fazer ou produzir. Assim, ao pensar, simultaneamente, produz-se e avalia-se, gera-se e julga-se os produtos que se constroem (Paul & Elder, 2005). Apenas o instrumento de Payan-Carreira et al. (2022) apresentaram um instrumento voltado para as capacidades de PC. Realça-se que nenhum dos autores inclui as três dimensões nas suas propostas. A Tabela 3 destaca os Instrumentos de Avaliação de PC, quanto às suas dimensões.

**Tabela 3.** Dimensões do PC (capacidades, disposições e hábitos mentais)

Autores	Instrumentos de Avaliação de PC		
	Disposições de PC	Capacidades de PC	Hábitos Mentais
Payan-Carreira, et al, (2022)		X	
Lopes, Silva & Morais (2021)	X		
Franco, et al. (2021)			X
Lopes, Silva & Morais (2019)		X*	
Franco, Costa & Almeida (2018)	X	X	

\*Pensamento criativo é associado ao PC

Fonte: dos autores

### 4.4. Instrumentos de Avaliação de PC – Base teórica de PC

De um modo geral, a construção de um instrumento de avaliação de PC tem por base uma conceitualização de PC (Facione & Facione, 1994). A base teórica mais escolhida foi a teoria de Facione, pelos seguintes autores: Payan-Carreira et al., (2022); Lopes et al. (2021); Franco et al. (2021). Destaca-se que os autores Lopes et al.(2019) usaram a taxonomia de Facione associada a outras duas taxonomias – de Bloom (revista) e de Guilford, esta especificidade

deve-se ao facto de este ser o único teste concebido de raiz. Apenas os autores Franco et al. (2018) seguiram a teoria de Halpern, como se verifica na tabela 4.

**Tabela 4.** Base teórica do PC

Autores	Instrumentos de Avaliação de PC
	Base teórica
Payan-Carreira, et al, (2022)	<i>Facione</i>
Lopes, Silva & Morais (2021)	<i>Facione</i>
Franco, et al. (2021)	<i>Facione</i>
Lopes, Silva & Morais (2019)	<i>Taxonomia de Bloom revista;</i> <i>Taxonomia de Facione (1990b);</i> <i>Taxonomia de Guilford (1986).</i>
Franco, Costa & Almeida (2018)	<i>Halpern</i>

Fonte: dos autores

#### 4.5. Níveis de ensino/Idades e tipos de cenários

Todos os Instrumentos de Avaliação do PC destinaram-se a estudantes do ES os instrumentos de Lopes et al. (2019) e Franco et al. (2018) estenderam-se a alunos de 17 anos e de 15 anos, respetivamente. Apenas os instrumentos de Lopes et al. (2019) e de Franco et al. (2018) apresentaram cenários da vida quotidiana. O primeiro apresentou uma situação-problema atual e comum do quotidiano, na qual se pode identificar um conjunto de circunstâncias problemáticas e um conjunto de seis perguntas, as quais se referem a diferentes tarefas cognitivas associadas ao domínio do PC correspondentes às capacidades de PC (Lopes et al., 2019). Estes dados apresentam-se na Tabela 5.

A ênfase dada ao ES corresponde à conclusão de que a maioria dos Instrumentos de Avaliação do PC destinam-se a adultos ou a estudantes do ES, com poucas adaptações para alunos mais jovens. Excetuam-se as versões dos testes Cornell (X, Y e Z), adaptáveis às faixas etárias, apesar de ainda carecem de maior consolidação psicométrica (Manassero-Mas & Vázquez-Alonso, 2024). A apresentação de cenários confere uma dimensão mais concreta ao instrumento, mais próxima da prática, a relevância destes cenários é destacada com a seguinte descrição: “O teste HCTA é considerado um ponto de viragem no cenário de avaliação do PC, principalmente pelo seu formato, pela apresentação de cenários da vida real” (Franco et al., 2018, p. 293).

**Tabela 5.** Níveis de ensino/Idades e tipos de cenários dos Instrumentos de Avaliação de PC

Autores	Instrumentos de Avaliação de PC	
	Níveis de Ensino/Idades	Tipos de cenários
Payan-Carreira, et al, (2022)	Ensino Superior	Sem cenários
Lopes, Silva & Morais (2021)	Ensino Superior	Sem cenários
Franco, et al. (2021)	Ensino Superior	Sem cenários
Lopes, Silva & Morais (2019)	Alunos a partir dos 17 anos Ensino Superior	1 cenário uma situação-problema atual e comum do quotidiano

Franco, Costa & Almeida (2018)	Alunos a partir dos 15 anos Ensino Superior	25 cenários da vida quotidiana de diversas áreas (educação, política, saúde, finanças, ...)
--------------------------------	--	---

Fonte: dos autores

#### 4.6. Instrumentos de Avaliação de PC – Tipos de Questões

Nos instrumentos de avaliação das capacidades de PC optaram-se por perguntas de resposta aberta longa ou mista, como é o caso dos autores Lopes et al. (2019) e Franco et al. (2018). Para os mesmos instrumentos os autores Payan-Carreira et al. (2022) deram preferência a perguntas de resposta fechada. Nos instrumentos de disposições de PC, Lopes, et al. (2021) usaram as escalas de Likert e Franco et al. (2018) usaram perguntas de resposta mista. Para os instrumentos de hábitos mentais, os autores Franco et al. (2021) optaram por perguntas de resposta fechada. Estas características encontram-se sintetizadas na tabela 6.

A diversidade do formato de perguntas corresponde à apresentada na literatura, com as vantagens e desvantagens inerentes a cada formato.

**Tabela 6.** Tipos de questões dos Instrumentos de Avaliação de PC

Autores	Instrumentos de Avaliação de PC				
	Tipos de questões				
	Aberta		Fechada		Mista
	Longa	Curta	Múltipla escolha	Escala Likert	
Payan-Carreira, et al. (2022)				X	
Lopes, Silva & Morais (2021)				X	
Franco, et al. (2021)			X		
Lopes, Silva & Morais (2019)	X				
Franco, Costa & Almeida (2018)					X

Fonte: dos autores

#### 4.7. Instrumentos de Avaliação de PC – Forças e Limitações

Relativamente aos instrumentos de avaliação das capacidades de PC, o instrumento de Lopes et al. (2019), por um lado, tinha como forças: uma fiabilidade altamente significativa; a especificidade de ter sido desenvolvido para o contexto português, com rigor metodológico e validade local. Por outro lado, tinha como limitação uma aplicabilidade limitada em contextos fora de Portugal. O instrumento de Franco et al. (2018) tinha como forças o facto de ter sido traduzido e adaptado para o português, porém sem a validação transcultural. Este tinha como limitação uma confiabilidade aceitável. Ambos os instrumentos tinham como mais-valia a existência de cenários da vida real.

Referente aos instrumentos de avaliação de disposições, o instrumento de Payan-Carreira et al. (2022) apresentou como força a aplicabilidade em contextos educacionais amplos e como limitação a validação limitada ao contexto português devido à falta de ajustes culturais. O instrumento de Lopes et al. (2021) apresentou uma boa fiabilidade como força, além da sua aplicabilidade em contextos educacionais amplos. Apresentou, também, como limitação uma aplicabilidade limitada em contextos fora de Portugal.

Por fim, relativo aos instrumentos de avaliação de hábitos mentais, o instrumento de Franco et al. (2021), o único identificado, apresentou como força a complementaridade deste instrumento aos seus anteriores voltados para as capacidades e para as disposições de PC. Este apresentou como limitação a ausência de confiabilidade ou fiabilidade.

#### 4.8. Instrumentos de Avaliação de PC – fundamentos para a seleção

Tendo por base as características dos instrumentos analisadas anteriormente, tal como as suas forças e limitações avançou-se para a definição dos fundamentos essenciais a ter em conta na seleção do instrumento de avaliação mais adequado, para uma tomada de decisão refletida. Os fundamentos definidos foram, de: abrangência; adequação cultural; validação; desenho de cenário; natureza de investigação; generalismo; extensão.

Sobre a abrangência dos instrumentos, destacaram-se por terem maior abrangência os instrumentos de Lopes et al. (2019; 2021) e de Franco et al. (2018; 2021). Os primeiros destacaram-se por avaliarem, além das capacidades, as disposições de PC, tendo a mesma base teórica relativamente ao PC. Esta base é considerada ampla dado que há a integração de várias teorias para conceberem instrumentos originais. O instrumento de avaliação das capacidades de PC, também, ganhou notoriedade por integrar o pensamento criativo. Ambos os pensamentos são inseparáveis, ao pensar, simultaneamente, produz-se e avalia-se, gera-se e julga-se os produtos que se constroem (Paul & Elder, 2005). Esta inclusão do pensamento criativo aumenta a aplicabilidade prática e a abrangência pedagógica do instrumento. Franco et al. (2018; 2021) destacaram-se por avaliar as capacidades, as disposições e, ainda, os hábitos mentais, garantindo uma avaliação mais completa e contextualizada do PC, apesar de usarem base teóricas diferentes. Tais bases não são integradas em cada instrumento, sendo que uns instrumentos baseiam-se numa teoria e outros noutras. Por outro lado, não separaram a dimensão das capacidades das disposições, os dois instrumentos estão integrados, “pelo que ambos os componentes estão fortemente conectados e por isso não são avaliados separadamente” (Franco et al., 2018, p. 294).

Relativamente à adequação cultural, os instrumentos de avaliação originais não requerem esta adequação dado que são concebidos atendendo às características dos potenciais utilizadores, como é o caso dos instrumentos de Lopes et al. (2019; 2021). A adequação cultural é essencial no caso dos instrumentos traduzidos, como é o caso do instrumento de Franco, et al. (2018), sendo que outro instrumento de Franco et al. (2021) ainda foi submetido a uma validação transcultural.

A validação dos instrumentos é uma garantia a assegurar. Entre os instrumentos de capacidades de PC destacou-se o teste de Payan-Carreira et al. (2022), com elevada confiabilidade, contudo trata-se um instrumento apenas traduzido, sem adequação cultural. Os outros instrumentos apresentaram: uma confiabilidade aceitável (Franco et al., 2018) e uma fiabilidade altamente significativa (Lopes et al., 2019). Dos instrumentos de disposições de PC não se destacaram tanto, o instrumento de Franco et al. (2018) apresentou uma confiabilidade aceitável e o de Lopes et al. (2021) apresentou uma fiabilidade boa. Por fim, o instrumento relativo aos hábitos mentais não foi validado.

Sobre o desenho de cenário da vida quotidiana destacaram-se unicamente os instrumentos de Lopes et al. (2019) e de Franco et al. (2018). O primeiro destaca-se por apresentar um único cenário permitindo um aprofundamento maior da situação-problema.

Quanto à natureza da investigação é relevante o tipo de questões dos instrumentos de avaliação. Para estudos com ênfase na natureza qualitativa é mais relevante os dados recolhidos por instrumentos de resposta aberta ou de resposta mista, neste sentido destacam-se instrumentos de capacidades de resposta aberta (Lopes et al., 2019) e os de resposta mista (Franco et al., 2018). Entre os instrumentos de disposições destaca-se o instrumento de resposta mista de Franco et al. (2018), embora não tenha ficado claro como o mesmo teste integra capacidades com disposições e perguntas de resposta aberta com resposta fechada.

Sobre o generalismo dos instrumentos - sem dependência de conteúdos específicos de uma disciplina, é essencial para abordagens interdisciplinares, a ausência de conteúdo específico de alguma disciplina amplia a sua aplicabilidade. Todos os instrumentos são generalistas, pelo que foram definidos critérios de inclusão e de exclusão para o efeito: Instrumentos de



Avaliação gerais, sem a integração de conteúdos de qualquer disciplina - critério de inclusão; Instrumentos de Avaliação específicos, com a integração de conteúdos da área disciplinar STEM ou a disciplinas de ciências ou de desporto – critério de exclusão.

A extensão é referente aos destinatários dos instrumentos. Todos os instrumentos são destinados a alunos, pelo que os critérios de inclusão e de exclusão assim restringiram os estudos. Os critérios definidos foram: Instrumentos de Avaliação de PC concebidos para alunos desde crianças a adultos – critério de inclusão; Instrumentos de Avaliação usados para avaliar alunos na transição para a vida adulta ou adultos em contexto de trabalho – Critério de Exclusão. Destacou-se o instrumento de Franco, Costa & Almeida (2018) por se destinar a alunos do ES e alunos ainda mais jovens – a partir dos 15 anos.

Selecionaram-se os instrumentos Lopes et al. (2019; 2021) atendendo ao seu destaque nos fundamentos de: abrangência, adequação cultural e desenho de cenário. Nestes instrumentos valorizamos: a sua originalidade, com a integração de diferentes teorias; a associação do pensamento criativo ao PC; a adequação cultural pensada de raiz; a apresentação de um único cenário permitindo um aprofundamento maior da situação-problema.

Selecionaria-se, também, os instrumentos de Franco et al. (2018) atendendo ao seu destaque no fundamento da extensão. Nestes instrumentos valorizou-se o facto de a extensão da idade permitir incluir destinatários do Ensino Secundário. Dada a necessidade de os instrumentos de Lopes et al. (2019; 2021) incluírem, igualmente, estes destinatários recorreu-se a outros meios para conhecer melhor tais instrumentos. Através da identificação de um livro, publicado em 2019, e do contacto direto com um dos autores. Da pesquisa adicional, encontrou-se no dito livro a extensão da aplicabilidade do instrumento de capacidades a alunos a partir dos 12 anos, além do acesso integral aos dois instrumentos, a grelhas de avaliação, a respostas, às cotações, ao tempo de aplicação e de correção.

## 5. CONCLUSÃO

Como referido na secção 3.1, a questão de investigação “Qual o modelo e instrumento de avaliação que integre as capacidades de PCC aplicável a alunos do secundário?”, neste artigo focou-se em identificar o modelo de pensamento crítico e criativo antes de incluir a dimensão da Inteligência Artificial Generativa através do ChatGPT. Assim, o objetivo desta RSL foi o de identificar e de selecionar na literatura um modelo e um instrumento de avaliação que integrasse as capacidades de PCC aplicável a alunos do secundário.

Com base nos critérios de inclusão e de exclusão definidos no protocolo da RSL, não foi identificado qualquer modelo de aprendizagem e foram identificados cinco instrumentos de avaliação. Destes instrumentos selecionamos os de Lopes et al. (2019; 2021).

Começando pelo Estudo de Natureza Qualitativa, o instrumento de Lopes et al. (2019) é especialmente indicado para estudos qualitativos, pois utilizou questões de resposta aberta, permitindo captar nuances do PCC dos alunos. Essas questões possibilitam análises aprofundadas das respostas dos participantes, alinhando-se perfeitamente aos objetivos de investigações qualitativas que buscam compreender processos e práticas educacionais complexas. O instrumento de Lopes et al. (2021) apesar de não usar perguntas de resposta aberta foi selecionado, dado que para a avaliação de disposições de PC não foi encontrado outro formato de pergunta. Sendo que este está interligado ao instrumento das capacidades de de Lopes et al. (2019), por ter a mesma base teórica e ter sido concebido, igualmente, de raiz, pelos mesmos autores.

Relativamente aos alunos do Ensino Secundário, o instrumento das capacidades de Lopes et al. (2019) destacou-se por ser concebido para o contexto do ensino secundário. Adicionalmente, a extensão identificada na publicação associada ao instrumento amplia sua aplicabilidade para estudantes a partir dos 12 anos. Essa flexibilidade permite atender às

necessidades de diferentes faixas etárias dentro do ensino secundário, promovendo uma avaliação adaptada ao nível cognitivo e educacional dos alunos. O instrumento das disposições de Lopes et al. (2021) continuou a ser aplicável a partir dos 17 anos.

Referente à Resolução de Problemas, o instrumento das capacidades de Lopes et al. (2019) inclui um cenário de resolução de problemas quotidianos, abordando questões práticas e relevantes para os estudantes. Essa abordagem permite avaliar como os alunos aplicam o pensamento crítico e criativo para solucionar problemas concretos, sendo uma característica essencial para medir capacidades que vão além da teoria e se aplicam à prática educacional e social.

Respeitante à Interdisciplinariedade, a integração do pensamento criativo com o PC, característica exclusiva do instrumento de Lopes et al. (2019), favorece uma abordagem interdisciplinar. A associação entre a criatividade e o PC é fundamental para resolver problemas que exigem conexões entre diferentes áreas do conhecimento, promovendo flexibilidade cognitiva e inovação, elementos centrais para abordar a complexidade dos problemas contemporâneos.

Por fim, no âmbito das Disciplinas das Ciências Sociais o mesmo instrumento é adequado por ter como foco a avaliação geral de capacidades críticas e criativas, sem dependência de conteúdos específicos de uma disciplina. A metodologia generalista e a relevância do cenário da vida quotidiana apresentadas garantem a sua aplicabilidade a temas interdisciplinares como educação, política e sociedade, promovendo uma análise integrada e contextualizada.

Como complemento, o instrumento de avaliação de hábitos mentais de Franco et al. (2021) poderá ser considerado, tornando a investigação mais abrangente, mantendo a garantia de uma base teórica coerente, atendendo a que se trata de uma adaptação de Facione (2015). O instrumento de avaliação selecionado, relativo às capacidades, será usado como matriz para o Modelo de Aprendizagem de PCC, ao qual será integrada com uma ferramenta de Inteligência Artificial Generativa, o ChatGPT, tirando-se partido das diferentes formas do seu uso e das especificidades de cada etapa do modelo e da respectiva capacidade em foco.

## AGRADECIMENTOS

O trabalho do primeiro autor é financiado por fundos nacionais. O trabalho do segundo autor é financiado por fundos nacionais.

## REFERÊNCIAS

- Al-massarweh, S. S., Khudairat, M. A., Freihat, R. H., & Jabali, S. M. (2024). The Impact of a Sterinberg Theory-Based Training Program on Secondary School Students' Development of Critical and Creative Thinking Skills. *International Journal of Religion*, 5(6), 438–456. <https://doi.org/10.61707/dp9g9y18>
- Bouckaert, M. (2023). The assessment of students' creative and critical thinking skills in higher education across OECD countries: A review of policies and related practices. *OECD Education Working Papers*, No. 293, 293. <https://doi.org/10.1787/35dbd439-en>
- Bradley, S., & Price, N. (2016). *CRITICAL THINKING Proven Strategies To Improve Decision Making Skills, Increase Intuition And Think Smarter!* (CreateSpace Independent Publishing Platform, Ed.).
- Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications. *Thinking Skills and Creativity*, 7(2), 112–121. <https://doi.org/10.1016/j.tsc.2012.04.001>
- Cardoso, T., Alarcão, I., & Celorico, J. (2010). *Revisão da Literatura e Sistematização do Conhecimento* (Porto Editora, Ed.).

- Choi, A. N., Flowers, S. K., & Heldenbrand, S. D. (2018). Becoming more holistic: A literature review of nonacademic factors in the admissions process of colleges and schools of pharmacy and other health professions. *Currents in Pharmacy Teaching and Learning*, 10(10), 1429–1437. <https://doi.org/10.1016/j.cptl.2018.07.013>
- Costa, A. P., Moreira, A., & Souza, F. N. de. (2019). webQDA - Qualitative Data Analysis (3.1). University of Aveiro and MicroIO. [www.webqda.net](http://www.webqda.net)
- Damopolii, I., Nunaki, J. H., Nusantara, E., & Kandowangko, N. Y. (2020). The effectiveness of Inquiry-based learning to train students' thinking skill based on SOLO taxonomy. *Journal of Physics: Conference Series*, 1567(4). <https://doi.org/10.1088/1742-6596/1567/4/042025>
- Danczak, S. M., Thompson, C. D., & Overton, T. L. (2020). Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills. *Chemistry Education Research and Practice*, 21(1), 62–78. <https://doi.org/10.1039/c8rp00130h>
- Davies, O., & Mansour, N. (2022). Exploring the Use of Cognitive Science Approaches Alongside SOLO Taxonomy as a Pedagogical Framework to Build Deeper Knowledge in Science and Foundation Subjects at Primary Schools in UK. *Education Sciences*, 12(8). <https://doi.org/10.3390/educsci12080523>
- Elder, L., & Paul, R. (2005). *Estándares de Competencia para el Pensamiento Crítico*. [www.criticalthinking.org](http://www.criticalthinking.org)
- Ennis, R. H. (2011). The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities.
- Facione, P. (1990). *The California Critical Thinking Skills Test: College Level Experimental Validation and Content Validity INSIGHT Reasoning Skills and Mindset measures for various professional groups View project Critical Thinking, Decision Making, and Problem Solving View project*. <https://www.researchgate.net/publication/268295703>
- Facione, N., & Facione, P. (1994). *The California Critical Thinking Skills Test and the li National League for Nursing ;Accreditation Requirement in Critical Thinking*.
- Facione, P. A. (2000). The Disposition Toward Critical Thinking: Its Character, Measurement, and Relationship to Critical Thinking Skill. *Informal Logic*, 20(1). <https://doi.org/10.22329/il.v20i1.2254>
- Facione, P. A., Facione, N. C., & Giancarlo, C. A. (2000). The Disposition Toward Critical Thinking: Its Character, Measurement, and Relationship to Critical Thinking Skill. *Informal Logic*, 20(1). <https://doi.org/10.22329/il.v20i1.2254>
- Facione, P. A. (2013). Critical Thinking: What It Is and Why It Counts. In *Measured Reasons*.
- Facione, P. A., Giancarlo, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. In *Journal of General Education* (Vol. 44, Issue 1).
- Facione, P., Gittens, C., & Facione, N. (2016). *Cultivating A Critical Thinking Mindset 1*. <https://www.researchgate.net/publication/311512219>
- Faravani, A., & Taleb, E. (2020). Teachers' use of Bloom's higher order questions in class to augment EFL learners' listening comprehension and critical thinking ability. *Journal of Language and Cultural Education*, 8(2), 94–113. <https://doi.org/10.2478/jolace-2020-0015>
- Franco, A. R., Costa, P. S., & Almeida, L. da S. (2018). Traducción, adaptación y validación del halpern critical thinking as-sessment en Portugal: Efecto del área disciplinaria y nivel académico en el pensamiento crítico. *Anales de Psicología*, 34(2), 292–298. <https://doi.org/10.6018/analesps.34.2.272401>
- Franco, A. R., Vieira, R. M., Riegel, F., & Crossetti, M. da G. O. (2021). Steering clear from 'lost in translation': cross-cultural translation, adaptation, and validation of critical thinking mindset self-rating form to university students. *Studies in Higher Education*, 46(3), 638–648. <https://doi.org/10.1080/03075079.2019.1647414>



- Gasmi, A. A., & al Nadabi, Z. S. (2023). An Exploratory Study of Learners' Perceptions About the Effectiveness of Active Learning Approaches. *Journal of Language Teaching and Research*, 14(5), 1223–1232. <https://doi.org/10.17507/jltr.1405.10>
- Hakim, N. W. A., & Talib, C. A. (2018). Measuring Critical Thinking in Science: Systematic Review. *Asian Social Science*, 14(11), 9. <https://doi.org/10.5539/ass.v14n11p9>
- Halpern, D. (1997). *Critical Thinking Across the Curriculum: A Brief Edition of Thought & Knowledge*. Routledge.
- Helaluddin, Mannahali, M., Purwati, D., Alamsyah, & Wijaya, H. (2023). An Investigation into the Effect of Problem-Based Learning on Learners' Writing Performance, Critical and Creative Thinking Skills. *Journal of Language and Education*, 9(2), 101–117. <https://doi.org/10.17323/jle.2023.14704>
- Lopes, J., Silva, H., & Morais, E. (2019). Critical and Creative Thinking Test for Higher Education Students. *Revista Lusofona de Educacao*, 44(44), 173–189. <https://doi.org/10.24140/issn.1645-7250.rle44.11>
- Lopes, J., Silva, H., & Morais, E. (2021). Development and validation of Critical Thinking Disposition Scale for college students (CTDS). *Revista Lusofona de Educacao*, 53(53), 119–138. <https://doi.org/10.24140/issn.1645-7250.rle53.07>
- Loyens, S. M. M., van Meerten, J. E., Schaap, L., & Wijnia, L. (2023). Situating Higher-Order, Critical, and Critical-Analytic Thinking in Problem- and Project-Based Learning Environments: A Systematic Review. In *Educational Psychology Review* (Vol. 35, Issue 2). Springer. <https://doi.org/10.1007/s10648-023-09757-x>
- Mas, M. A. M., & Alonso, Á. V. (2024). Making thinking skills visible in elementary education: psychometric development of an evaluation tool. *Bordon. Revista de Pedagogia*, 76(1), 119–139. <https://doi.org/10.13042/Bordon.2024.95702>
- Moons, P., Goossens, E., & Thompson, D. R. (2021, June 1). Rapid reviews: the pros and cons of an accelerated review process. *European Journal of Cardiovascular Nursing*, 20(5), 515–519. <https://doi.org/10.1093/eurjcn/zvab041>
- Nieto, A. M., & Valenzuela, J. (2012). A Study of the Internal Structure of Critical Thinking Dispositions. *Inquiry: Critical Thinking Across the Disciplines*, 27(1), 31–38. <https://doi.org/10.5840/inquiryct20122713>
- Ossa-Cornejo, C. J., Palma-Luengo, M. R., Lagos-San Martín, N. G., Quintana-Abello, I. M., & Díaz-Larenas, C. H. (2017). ANÁLISIS DE INSTRUMENTOS DE MEDICIÓN DEL PENSAMIENTO CRÍTICO. *Ciencias Psicológicas*, 19–28. <https://doi.org/10.22235/cp.v11i2.1343>
- Paula, A., Ferraz, C. M., & Belhot, R. V. (2010). Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais Bloom's taxonomy and its adequacy to define instructional objective in order to obtain excellence in teaching. *Canadá. Gest. Prod*, 2, 421–431.
- Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). *Development and Validation of a Critical Thinking Assessment-Scale Short Form*. <https://doi.org/10.3390/educsci>
- Quinn, S., Hogan, M., Dwyer, C., Finn, P., & Fogarty, E. (2020). Development and Validation of the Student-Educator Negotiated Critical Thinking Dispositions Scale (SENCTDS). *Thinking Skills and Creativity*, 38. <https://doi.org/10.1016/j.tsc.2020.100710>
- Rear, D. (2019). One size fits all? The limitations of standardised assessment in critical thinking. *Assessment and Evaluation in Higher Education*, 44(5), 664–675. <https://doi.org/10.1080/02602938.2018.1526255>
- Ross, D., Loeffler, K., Schipper, S., Vandermeer, B., & Allan, G. M. (2013). Do scores on three commonly used measures of critical thinking correlate with academic success of health professions

trainees? A systematic review and meta-analysis. *Academic Medicine*, 88(5), 724–734.  
<https://doi.org/10.1097/ACM.0b013e31828b0823>

Saiz, C., & Rivas, S. F. (2008). *INTERVENIR PARA TRANSFERIR EN PENSAMIENTO CRÍTICO*.  
<http://www.pensamiento-critico.com/pensacono/autor.htm>

Schünemann, H. J., & Moja, L. (2015). Reviews: Rapid! Rapid! Rapid!.and systematic. *Systematic Reviews*, 4(1). <https://doi.org/10.1186/2046-4053-4-4>

Soares, C., Fornari, L., Pinho, I., & Costa, A. (2019). *Revisão de Literatura com Apoio de Software: Contribuição Qualitativa*. Ludomedia. [www.ludomedia.org](http://www.ludomedia.org)-

Tenreiro-Vieira, C., & Vieira, R. M. (2022). Pensamento crítico e criativo para uma educação ciência-tecnologia-sociedade Pensamiento crítico y creativo para una educación ciencia-tecnología-sociedad Critical and Creative Thinking for Science-Technology-Society Education. *Revista CTS*, 17, 141–155.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=8738127>

Thamrin, L., Gustian, U., Zhongfulin, W., & Suryadi, D. (2024). The Implementation of Contextual Learning Strategies to Stimulate Students' Critical Thinking Skills La aplicación de estrategias de aprendizaje contextual para estimular la capacidad de pensamiento crítico de los estudiantes. *Retos Nuevas Tendencias En Educación Física, Deporte y Recreación*.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=9294849>

Yue, M., Zhang, M., Zhang, C., & Jin, C. (2017). The effectiveness of concept mapping on development of critical thinking in nursing education: A systematic review and meta-analysis. In *Nurse Education Today* (Vol. 52, pp. 87–94). Churchill Livingstone.  
<https://doi.org/10.1016/j.nedt.2017.02.018>

Zapalska, A. M., McCarty, M. D., Young-McLear, K., & White, J. (2018). Design of assignments using the 21st century Bloom's revised taxonomy model for development of critical thinking skills. *Problems and Perspectives in Management*, 16(2), 291–305.  
[https://doi.org/10.21511/ppm.16\(2\).2018.27](https://doi.org/10.21511/ppm.16(2).2018.27)