Received: 16 December 2024 | Accepted: 18 February 2026 | Published: 13 March 2026

ARK: 24285/RCC.v10i17.306

ARTICLE

REVISTA COLETA CIENTÍFICA

Comparabilidade e equidade na avaliação da matemática no PISA 2022 e no TIMSS 2023

Comparability and equity in mathematics assessment in PISA 2022 and TIMSS 2023

Raphael Alves de Oliveira[1]
https://orcid.org/0009-0008-6608-3298
https://lattes.cnpq.br/9512157066307239
Universidade de Brasília (UnB), Distrito Federal, Brasil
E-mail: raphalvez@gmail.com

Informações da publicação

ARK: 24285/RCC.v10i17.306

ISSN: 2763-6496

Palavras-chave:
Educação comparada.
Avaliação educacional.
Comparabilidade.
Equidade.
Matemática.

Keywords:
Comparative education.
Educational assessment.
Comparability.
Equity.
Mathematics.

Abstract:
International assessments provide information for educational research, but their results depend on choices concerning populations, knowledge and context. This article examines the conditions for comparability between the 2022 Programme for International Student Assessment and the 2023 Trends in International Mathematics and Science Study, focusing on mathematics. It presents a qualitative documentary study based on a directed reading of official frameworks and a contrast across six analytical categories: purpose, population, construct, knowledge organization, contextual information and scope of inference. The discussion draws on comparative education, the validity of score interpretations and educational equity. The documentary contrast identifies shared attention to reasoning and problem solving alongside differences in student eligibility and domain organization. These differences require researchers to distinguish comparisons of assessment designs from equivalence between results. The article proposes a set of questions to make the conditions and limitations of research using international assessments explicit. It concludes that equity should inform the construction of the comparative object, particularly regarding the population represented and the learning opportunities considered. The proposed framework is interpretive and has not been empirically validated; it does not provide a method for converting scores across programmes.

[1] Graduado em Educação Física pela Universidade de Brasília. Graduado em Pedagogia pela Faculdade Aberta do Tocantins. Pós-graduado em Educação Infantil e Alfabetização pela Faculdade Catedral.

REVISTA
COLETA CIENTÍFICA

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

Resumo:
As avaliações internacionais oferecem informações para a pesquisa educacional, mas seus resultados dependem de escolhas sobre população, conhecimento e contexto. Este artigo analisa as condições de comparabilidade entre o Programme for International Student Assessment, edição de 2022, e o Trends in International Mathematics and Science Study, edição de 2023, com delimitação no domínio da matemática. Desenvolve-se uma pesquisa qualitativa documental, baseada na leitura dirigida dos referenciais oficiais dos dois programas e no contraste entre seis categorias analíticas: finalidade, população, construto, organização do conhecimento, informações contextuais e alcance das inferências. A discussão articula contribuições da educação comparada, da validade das interpretações de escores e da equidade educacional. O contraste documental identifica aproximações no interesse pelo raciocínio e pela resolução de problemas, acompanhadas de diferenças na delimitação dos estudantes e na organização dos domínios avaliados. Argumenta-se que essas diferenças exigem distinguir a comparação dos desenhos avaliativos da equivalência entre resultados. Como contribuição, apresenta-se um roteiro de perguntas para explicitar condições e limites de estudos que utilizem avaliações internacionais. Conclui-se que a equidade deve integrar a construção do objeto comparativo, particularmente quanto à população representada e às oportunidades de aprendizagem consideradas. O roteiro constitui uma proposta interpretativa, sem validação empírica ou pretensão de converter escores entre programas.

1 Introdução

Comparar resultados educacionais exige definir o que será tratado como semelhante e quais diferenças permanecerão relevantes. Quando duas avaliações utilizam a expressão desempenho em matemática, a proximidade do vocabulário pode sugerir que seus resultados descrevem o mesmo objeto. Entretanto, o significado de um resultado depende de quem participa da avaliação, das tarefas propostas e do conhecimento que se pretende representar. A investigação comparativa começa, portanto, antes da leitura das pontuações.

No debate sobre avaliações internacionais, Silveira e Esquinsani (2024) examinam a educação comparada em sua relação com organismos internacionais e com políticas de avaliação externa. O presente estudo dialoga com essa temática por outra entrada: investiga as condições metodológicas necessárias para aproximar informações produzidas por programas distintos. O objeto deixa de ser a trajetória de determinada política e passa a ser a construção da comparabilidade entre desenhos avaliativos.

São considerados o Programme for International Student Assessment (PISA), coordenado pela Organização para a Cooperação e Desenvolvimento Econômico, e o Trends in International Mathematics and Science Study (TIMSS), desenvolvido no âmbito da International Association for the Evaluation of Educational Achievement. A seleção recai sobre matemática no PISA 2022 e no TIMSS 2023. Trata-se de um recorte documental delimitado, e não da afirmação de que esses sejam os ciclos mais recentes dos programas.

A questão que orienta o artigo é: em que condições os referenciais dessas avaliações permitem construir comparações educacionalmente válidas sem apagar diferenças entre as populações e as oportunidades de aprendizagem? O objetivo geral é analisar as aproximações e os limites de comparabilidade dos dois desenhos. Especificamente, busca-se explicitar os critérios de contraste, discutir a presença do contexto na interpretação e propor perguntas que auxiliem a delimitação de futuras pesquisas.

A tese desenvolvida é que a comparabilidade precisa ser justificada em relação à pergunta de pesquisa. Dois programas podem ser comparados quanto às suas concepções de aprendizagem sem que suas pontuações sejam intercambiáveis. De modo semelhante,

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

a possibilidade de descrever uma associação entre condições sociais e desempenho não autoriza, por si, explicações causais sobre escolas ou estudantes. Distinguir essas operações torna o argumento científico mais preciso.

A contribuição pretendida é metodológica. O estudo apresenta um roteiro interpretativo que relaciona a definição do objeto à equidade na representação dos estudantes. Não se calculam resultados nacionais, não se estabelece uma classificação de sistemas educacionais e não se propõe a conversão entre escalas. O que se examina é a consistência das perguntas formuladas com o tipo de evidência que os documentos permitem reconhecer.

2 Educação comparada e validade das interpretações

A comparação educacional pode investigar instituições, processos de aprendizagem e formas de produzir conhecimento sobre a escolarização. A seleção do país como unidade, embora frequente em avaliações internacionais, não encerra as possibilidades do campo. Bray e Thomas (1995) defendem a ampliação dos níveis de comparação e a articulação de perspectivas que, quando isoladas, oferecem leituras parciais. Essa contribuição permite questionar o uso de uma média nacional como descrição suficiente da experiência escolar.

Bartlett e Vavrus (2017) propõem uma abordagem de estudos de caso comparativos atenta às relações horizontais, verticais e transversais. Neste artigo, a referência contribui para reconhecer que contexto e escala precisam participar da construção do problema. Não se aplica integralmente seu desenho de estudo de caso, pois o material aqui examinado consiste em referenciais institucionais, sem investigação de campo ou acompanhamento de processos locais.

Essa delimitação importa porque uma análise documental pode identificar como um programa descreve seu objeto, mas não comprova que todas as situações de aplicação correspondam ao desenho anunciado. O documento fornece evidências sobre definições e intenções institucionais. A distância entre essas formulações e a experiência dos participantes requer outras fontes. Confundir os dois planos produziria uma conclusão maior do que a evidência disponível.

A noção de validade oferece um segundo fundamento. Nos Standards for Educational and Psychological Testing, a validade se refere ao apoio teórico e empírico às interpretações de escores para usos determinados (American Educational Research Association et al., 2014). Assim, a qualidade de um instrumento não torna automaticamente aceitável qualquer conclusão extraída de seus resultados. Uma interpretação precisa ser examinada em relação ao propósito que pretende atender.

Aplicado à comparação entre programas, esse princípio conduz à seguinte exigência: o pesquisador deve declarar qual relação pretende estabelecer. Comparar definições de raciocínio matemático é uma operação conceitual. Estimar diferenças de proficiência é uma operação estatística que depende de condições adicionais. Explicar efeitos de uma experiência escolar demanda um desenho capaz de sustentar essa explicação. Essas operações podem integrar uma agenda comum, mas requerem argumentos próprios.

Neste estudo, emprega-se comparabilidade documental para designar a possibilidade de contrastar definições e critérios de organização explicitados nos referenciais. Equivalência de medida designa uma questão distinta: se os resultados

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

podem representar uma grandeza comum com significado suficientemente estável para o uso pretendido. O primeiro tipo de exame auxilia a formular o segundo, mas não o resolve. As expressões funcionam como distinções operacionais do artigo, e não como uma nova classificação psicométrica validada.

A equidade acrescenta uma pergunta sobre quem está presente na evidência. O guia da UNESCO (2017) relaciona inclusão e equidade à identificação de barreiras que afetam participação e aprendizagem. Adota-se essa orientação como referência conceitual para examinar a construção do objeto comparativo. Não se utiliza o documento para atribuir uma condição de justiça ou injustiça a qualquer programa específico.

Sob essa perspectiva, a população representada e as condições de aprendizagem são partes do significado dos resultados. Um estudo que se refira genericamente aos jovens de um território precisa demonstrar que sua evidência alcança esse conjunto. Se os dados dizem respeito apenas aos matriculados em determinadas condições, a redação deve conservar essa restrição. A equidade participa, nesse caso, da precisão do conhecimento produzido.

3 Percurso metodológico

O estudo tem abordagem qualitativa e delineamento documental comparativo. As unidades de análise são os desenhos de avaliação da matemática no PISA 2022 e no TIMSS 2023, tal como apresentados em seus referenciais oficiais. A comparação não toma estudantes, escolas ou países como casos empíricos. Seu nível de observação é o da documentação que define o objeto avaliado.

A seleção do corpus foi intencional e orientada por três critérios: autoria institucional identificável, acesso público ao texto e relação direta com a definição da avaliação ou de suas informações contextuais. No PISA, foram examinados os capítulos de apresentação geral, matemática e questionários do PISA 2022 Assessment and Analytical Framework. No TIMSS, foram considerados a introdução e os capítulos de matemática e questionários contextuais do TIMSS 2023 Assessment Frameworks. As seções do primeiro documento foram consultadas na edição de 2023; as do segundo, na edição de 2021. O ano do ciclo de aplicação não foi tratado como ano de publicação.

Como documentação complementar, consultaram-se o guia de leitura, o anexo sobre população e amostragem e as notas técnicas de PISA 2022 Results Volume I (OECD, 2023b). Esses materiais auxiliaram a delimitar a linguagem sobre representação e incerteza. Os textos de educação comparada, os Standards e o guia da UNESCO constituíram apoio conceitual, sem serem tratados como documentos equivalentes aos referenciais dos programas.

Quadro 1

Documentos nucleares e função na análise

Unidade documental	Referência	Aspecto examinado
Apresentação e matemática do PISA 2022	OECD (2023a), capítulos 1 e 2	Finalidade, população e definição do domínio
Questionários do PISA 2022	OECD (2023a), capítulo 5	Dimensões contextuais previstas

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

Unidade documental	Referência	Aspecto examinado
Introdução e matemática do TIMSS 2023	Mullis et al. (2021); Philpot et al. (2021)	Vínculo curricular e organização do domínio
Questionários contextuais do TIMSS 2023	Reynolds et al. (2021)	Informações sobre ambientes de aprendizagem

Fonte: organização da análise a partir dos documentos indicados.

A leitura foi dirigida por seis categorias definidas em função do problema: finalidade declarada, população de referência, construto avaliado, organização do conhecimento, informações contextuais e alcance das inferências. Foram extraídas definições e descrições pertinentes a essas categorias. Em seguida, os registros foram colocados em contraste, conservando a identificação do documento e da seção de origem. As denominações em inglês foram mantidas nas referências e vertidas para o português no desenvolvimento do texto.

A análise distinguiu três níveis de enunciado. O primeiro registra uma característica expressamente apresentada no referencial. O segundo formula uma interpretação comparativa dessa característica. O terceiro apresenta uma proposta de organização para pesquisas futuras. Essa distinção evita que uma recomendação do artigo seja atribuída às instituições responsáveis pelas avaliações. Os quadros de síntese incluem formulações analíticas próprias e não reproduzem tabelas dos documentos.

Não foi realizada revisão sistemática da literatura, aplicação de instrumento a participantes, codificação independente por avaliadores ou estimação com microdados. O procedimento consiste em leitura analítica e contraste argumentativo. Sua rastreabilidade decorre da delimitação do corpus, da identificação das categorias e dos localizadores documentais. As conclusões se restringem à configuração descrita nas fontes selecionadas; não avaliam a qualidade empírica de cada aplicação.

4 Aproximações e diferenças entre os desenhos avaliativos

4.1 População e significado da comparação

O PISA utiliza uma definição etária de população e contempla estudantes de aproximadamente 15 anos matriculados a partir do equivalente ao sétimo ano de escolaridade, segundo os critérios detalhados no anexo de amostragem (OECD, 2023b). O TIMSS se organiza em torno do quarto e do oitavo anos de escolaridade, conforme sua introdução (Mullis et al., 2021). Esses critérios definem referências diferentes para interpretar a participação.

A consequência analítica é que idade e escolaridade não podem ser tratadas como variáveis intercambiáveis. A delimitação etária permite investigar uma etapa da vida escolar, preservando diferenças no percurso dos participantes. A delimitação por ano de escolaridade toma uma posição no percurso educacional como referência. Em ambos os casos, é necessário formular a população sobre a qual se pretende falar, sem substituir a definição efetiva por uma expressão mais abrangente.

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

Considere-se uma situação hipotética em que dois grupos de mesma idade apresentem percursos escolares diferentes. Uma comparação baseada apenas em idade conservará essa diversidade; outra, organizada por ano cursado, responderá a uma pergunta diferente. Não se deduz daí a superioridade de uma alternativa. O exemplo evidencia que a escolha do critério modifica o objeto descrito. Tampouco se afirma que os documentos examinados mensurem de forma idêntica todos os componentes desses percursos.

No âmbito da equidade, essa diferença suscita o exame da cobertura. Os estudantes efetivamente representados precisam ser identificados antes de se falar na experiência educacional de uma geração inteira. O pertencimento a uma faixa etária não torna automaticamente todos os seus integrantes elegíveis para uma avaliação escolar. A condição de matrícula integra a delimitação. Assim, uma conclusão sobre participantes elegíveis não deve se transformar, por alteração de linguagem, em conclusão sobre todos os jovens.

4.2 Construto e organização do conhecimento matemático

No PISA 2022, o letramento matemático articula raciocínio e processos de formular, empregar e interpretar matemática em problemas situados. O conteúdo se organiza em mudança e relações, espaço e forma, quantidade e incerteza e dados (OECD, 2023a). No TIMSS 2023, os domínios cognitivos são conhecer, aplicar e raciocinar; o documento também contempla resolução de problemas em contexto (Philpot et al., 2021). A proximidade entre algumas expressões não estabelece identidade entre os construtos.

A leitura comparativa permite reconhecer que a oposição entre uma avaliação de aplicação e outra de memorização seria insuficiente para descrever os referenciais. O contraste relevante está no modo como cada desenho articula seus componentes. Para investigar essa articulação, é necessário examinar as relações entre conteúdos, processos e tarefas, em vez de transformar o título de um domínio em uma definição exaustiva do programa.

Essa observação altera a unidade mínima de uma futura análise de itens. O termo raciocínio, isoladamente, pode designar operações com exigências distintas. Uma comparação precisaria considerar o enunciado, os recursos disponíveis, a resposta solicitada e os critérios de correção. Dois itens sobre o mesmo conteúdo podem demandar operações diferentes; itens situados em categorias diferentes podem compartilhar parte das exigências. O presente estudo formula essa possibilidade como problema de investigação, sem apresentar uma codificação de itens que a demonstre.

Um exemplo elaborado para esta discussão ajuda a precisar a diferença. Imagine-se uma tarefa que informe a capacidade de dois reservatórios e solicite uma diferença de volumes. Outra tarefa utiliza os mesmos valores, mas pede que o estudante escolha uma proposta de abastecimento e justifique sua escolha sob uma restrição de consumo. O conteúdo numérico pode ser semelhante, enquanto a segunda tarefa acrescenta decisões sobre o modelo e sua interpretação. Inversamente, mudar o cenário de abastecimento para transporte não demonstra, por si, uma alteração da operação matemática exigida. O exemplo não reproduz itens dos programas e não pretende classificá-los; mostra por que conteúdo, contexto e demanda cognitiva precisam ser examinados separadamente.

Também é necessário distinguir semelhança temática de equivalência de pontuação. Identificar conteúdos comuns serve à construção de perguntas sobre o que os

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

instrumentos privilegiam. Não fornece uma transformação matemática entre escalas. Qualquer estudo que pretendesse estabelecer essa transformação precisaria de dados e procedimentos próprios de vinculação, acompanhados de justificativas para o uso proposto. A análise documental realizada não produz esse tipo de evidência.

Quadro 2
Síntese do contraste documental e de suas implicações

Dimensão	Característica identificada	Implicação analítica proposta
População	Referência etária no PISA e por escolaridade no TIMSS	Declarar o universo de cada conclusão
Conhecimento	Letramento matemático e organização curricular com ênfases distintas	Examinar o construto antes de aproximar escores
Processos	Presença de raciocínio e resolução de problemas em ambos	Não reduzir o contraste a aplicação versus memorização
Contexto	Previsão de informações além das respostas cognitivas	Identificar o que cada variável representa
Inferência	Documentos definem desenhos e critérios de interpretação	Separar descrição documental, estimação e explicação causal

Fonte: síntese interpretativa baseada em OECD (2023a, 2023b), Mullis et al. (2021), Philpot et al. (2021) e Reynolds et al. (2021). As implicações são proposições deste artigo.

4.3 Contexto e oportunidades de aprendizagem

O referencial dos questionários do PISA prevê informações contextuais, enquanto o do TIMSS contempla ambientes familiares, escolares e de sala de aula, atributos dos estudantes e contexto nacional (OECD, 2023a; Reynolds et al., 2021). No segundo, recursos domésticos e experiências de ensino figuram entre as dimensões examinadas. Essas descrições indicam que os programas não se limitam à apresentação de respostas cognitivas.

A existência de questionários, contudo, não resolve a interpretação das desigualdades. Para uma variável cumprir função analítica, o pesquisador precisa explicitar o que ela representa e quais aspectos deixa sem observação. Um indicador de recursos materiais pode aproximar determinadas condições de estudo, mas não constitui descrição completa da vida familiar. Sua relação com a aprendizagem deve ser investigada de acordo com a definição adotada, evitando a substituição de pessoas e experiências por um único indicador.

Propõe-se distinguir contexto como característica da população e contexto como oportunidade de aprendizagem. A primeira perspectiva descreve condições em que os participantes se encontram. A segunda pergunta pelas experiências educacionais às quais tiveram acesso. As perspectivas podem se relacionar, mas não são equivalentes. Uma análise que as reúna sem justificativa corre o risco de converter uma característica social em explicação automática do desempenho.

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

O mesmo cuidado se aplica à comparação internacional de categorias contextuais. Uma denominação comum precisa ser examinada em relação ao modo de coleta e ao significado local. A disponibilidade de um recurso pode não descrever a frequência, a finalidade ou a qualidade de seu uso. Neste artigo, esse ponto constitui uma exigência interpretativa para pesquisas futuras. Não se apresenta evidência de falha específica de tradução, de mensuração ou de equivalência em qualquer instrumento analisado.

4.4 Incerteza e alcance das inferências

O guia de leitura e as notas técnicas do PISA 2022 explicitam procedimentos de tratamento da incerteza e das estimativas de desempenho (OECD, 2023b). Para o argumento deste artigo, a consequência é circunscrita: a diferença observada entre estimativas requer interpretação compatível com a precisão da evidência. Uma descrição pontual não informa, sozinha, a estabilidade ou a relevância educacional de um contraste.

Há ainda uma diferença lógica entre identificar associação e explicar um mecanismo. Se duas características aparecem relacionadas, essa observação pode orientar hipóteses, mas não estabelece automaticamente que uma produziu a outra. Uma explicação sobre práticas escolares precisa considerar como as informações foram produzidas e que alternativas explicativas permanecem possíveis. O contraste de referenciais não testa essas hipóteses e, por isso, não atribui efeitos às características descritas.

Essa limitação tem valor positivo para a pesquisa: permite determinar o que precisa ser investigado em uma etapa posterior. Uma divergência entre leituras de avaliações pode suscitar perguntas sobre população, currículo ou condições de participação. Ela não exige que se escolha imediatamente um programa como representação definitiva da aprendizagem. O tratamento da divergência como problema, em vez de conclusão pronta, amplia as possibilidades de investigação.

5 Um roteiro para explicitar as condições da comparação

Com base no contraste desenvolvido, propõe-se um roteiro de seis perguntas para estudos que utilizem avaliações internacionais. Sua função é organizar a justificativa metodológica e tornar visível o percurso entre pergunta, evidência e conclusão. Trata-se de uma elaboração interpretativa deste artigo. Não é uma escala, não atribui notas aos programas e não substitui análise especializada de mensuração.

Quadro 3

Perguntas para delimitar estudos comparativos de avaliação

Critério	Pergunta orientadora	Registro esperado no estudo
Objeto	O que exatamente se pretende comparar	Definição da aprendizagem ou do processo investigado
População	Quem a evidência representa e quem fica fora dela	Elegibilidade, cobertura e limites de generalização
Conhecimento	Que operações e conteúdos sustentam a aproximação	Correspondências e diferenças justificadas

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

Critério	Pergunta orientadora	Registro esperado no estudo
Oportunidades	Que experiências de aprendizagem entram na interpretação	Indicadores escolhidos e alcance de seus significados
Inferência	A conclusão descreve, estima, associa ou explica	Compatibilidade entre desenho e tipo de afirmação
Contexto	Que informação adicional pode alterar a leitura inicial	Fontes complementares e questões ainda abertas

Fonte: elaboração própria em diálogo com os referenciais analisados, American Educational Research Association et al. (2014), Bray e Thomas (1995), Bartlett e Vavrus (2017) e UNESCO (2017).

A aplicação começa pela definição do objeto. Se o interesse estiver nas concepções de aprendizagem, o estudo pode trabalhar com matrizes e tarefas divulgadas. Se estiver em resultados de populações, precisará incorporar amostragem e estimação. Se a questão for compreender experiências de ensino, dados contextuais podem precisar ser articulados a outras fontes. O roteiro explicita essas diferenças sem estabelecer uma sequência universal de métodos.

A segunda e a quarta perguntas conectam representação e equidade. Identificar quem está fora da evidência e quais oportunidades são consideradas impede que os limites de observação permaneçam ocultos. Também auxilia a redigir conclusões proporcionais ao estudo. A finalidade não é explicar antecipadamente qualquer diferença por uma condição social, mas declarar quais aspectos foram observados e quais exigem investigação adicional.

A terceira pergunta organiza a comparação conceitual. Em um estudo sobre matemática, isso significa demonstrar por que duas descrições ou tarefas podem ser aproximadas. Uma resposta satisfatória pode reconhecer correspondências parciais e conservar diferenças relevantes. A comparação não exige identidade completa dos objetos; exige que a relação proposta seja argumentada. Esse princípio permite produzir conhecimento mesmo quando a equivalência não pode ser sustentada.

A quinta pergunta estabelece um controle sobre a linguagem da conclusão. Verbos como descrever, associar e explicar expressam compromissos diferentes. O pesquisador pode revisar cada afirmação perguntando qual evidência a sustenta. Caso a conclusão ultrapasse o desenho utilizado, deve ser reformulada como hipótese ou como questão para pesquisa posterior. O procedimento não depende de uma classificação de qualidade dos programas.

Por fim, a pergunta sobre contexto mantém a análise aberta a informações que contrariem a interpretação inicial. Documentos curriculares, registros institucionais e observações de práticas podem esclarecer aspectos que uma avaliação padronizada não foi concebida para descrever. A necessidade dessas fontes depende do problema investigado. Sua inclusão exige planejamento próprio, e não pode ser simulada por inferências sobre experiências que não foram observadas.

O roteiro apresenta limites. As categorias são amplas e ainda não foram testadas em um conjunto de estudos independentes. A aplicação a áreas diferentes da matemática pode requerer alterações. Também não se estabeleceu concordância entre avaliadores na classificação dos documentos. Uma agenda de continuidade poderia examinar sua

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

utilidade em análises de itens divulgados e em investigações que articulem resultados e experiências escolares, preservando as exigências específicas de cada desenho.

6 Considerações finais

A análise documental respondeu à questão de pesquisa ao identificar que a comparação entre PISA 2022 e TIMSS 2023 exige conservar diferenças de população, construto e organização do conhecimento. As aproximações no interesse pelo raciocínio e pela resolução de problemas oferecem um terreno para investigação conceitual. Elas não demonstram equivalência entre escalas nem autorizam a transferência direta de conclusões de um programa para outro.

A principal implicação metodológica consiste em distinguir a comparação dos desenhos da comparação dos resultados e da explicação das diferenças. Cada operação estabelece exigências próprias. O alcance das conclusões deve permanecer vinculado à evidência efetivamente examinada. No caso deste artigo, os achados se referem aos referenciais selecionados e às possibilidades interpretativas reconhecidas em sua leitura.

A equidade foi incorporada como pergunta sobre representação e oportunidades de aprendizagem. Isso conduz à explicitação de quem participa do universo observado e de quais experiências são consideradas na interpretação. O artigo não determina se um programa é mais equitativo que outro. Propõe que as condições de participação e o significado das variáveis contextuais integrem a justificativa de qualquer aproximação entre seus dados.

O roteiro apresentado sintetiza a contribuição do estudo. Suas perguntas podem auxiliar a formular pesquisas que conservem a complexidade dos objetos e indiquem com clareza o que ainda não foi estabelecido. Sua utilidade deverá ser examinada em aplicações posteriores. A comparação educacional, nessa perspectiva, produz conhecimento ao tornar justificáveis as relações entre objetos, sem exigir que diferenças de desenho sejam apagadas para que a análise possa ocorrer.

Referências

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Bartlett, L., & Vavrus, F. (2017). Comparative case studies: An innovative approach. *Nordic Journal of Comparative and International Education, 1*(1). https://doi.org/10.7577/njcie.1929

Bray, M., & Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. *Harvard Educational Review, 65*(3), 472–491. https://doi.org/10.17763/haer.65.3.g3228437224v4877

Mullis, I. V. S., Martin, M. O., & von Davier, M. (2021). Introduction. In I. V. S. Mullis, M. O. Martin, & M. von Davier (Eds.), *TIMSS 2023 assessment frameworks* (pp. 1–4). TIMSS & PIRLS International Study Center, Boston College.

© 2026 The authors
This article is licensed under the Creative Commons Attribution 4.0 International License — CC BY 4.0

https://timssandpirls.bc.edu/timss2023/frameworks/pdf/T23_Frameworks_Introduction.pdf

OECD. (2023a). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

OECD. (2023b). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Philpot, R., Lindquist, M., Mullis, I. V. S., & Aldrich, C. E. A. (2021). TIMSS 2023 mathematics framework. In I. V. S. Mullis, M. O. Martin, & M. von Davier (Eds.), *TIMSS 2023 assessment frameworks* (pp. 5–18). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2023/frameworks/pdf/T23_Frameworks_Ch1_Mathematics.pdf

Reynolds, K. A., Mullis, I. V. S., & Martin, M. O. (2021). TIMSS 2023 context questionnaire framework. In I. V. S. Mullis, M. O. Martin, & M. von Davier (Eds.), *TIMSS 2023 assessment frameworks* (pp. 46–70). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2023/frameworks/pdf/T23_Frameworks_Ch3_Context-Questionnaires.pdf

Silveira, C. L. A. da, & Esquinsani, R. S. S. (2024). A educação comparada como evidência para as políticas de avaliação externa em educação. *Cuadernos de Educación y Desarrollo, 16*(4), 1–26. https://doi.org/10.55905/cuadv16n4-005

UNESCO. (2017). *A guide for ensuring inclusion and equity in education*. https://unesdoc.unesco.org/ark:/48223/pf0000248254