

TESTES EDUCACIONAIS EM LARGA ESCALA: ASPECTOS POLÍTICOS, PEDAGÓGICOS E METODOLÓGICOS

LARGE SCALE EDUCATIONAL TESTS: POLITICAL, PEDAGOGICAL AND METHODOLOGICAL ASPECTS

Leandro Araujo de Sousa¹
Adriana Eufrásio Braga²

RESUMO

Este ensaio objetivou problematizar as possíveis implicações políticas, pedagógicas e metodológicas dos testes educacionais em larga escala. Para tanto, realizamos um levantamento dos testes educacionais nacionais e estaduais realizados atualmente no Brasil. Os achados indicam que os testes educacionais começaram a ser implantados no início da década de 1990, a priori utilizando de psicometria clássica. Ao longo dos anos os governos passaram a utilizar os modelos de teoria de resposta ao item unidimensionais na análise dos resultados, assim como os testes implantados mais recentemente. Esses modelos de análise dos resultados dos testes têm sido amplamente adotados no Brasil. No entanto, questiona-se a adequação dos testes aos pressupostos do modelo estatístico de teoria de resposta ao item unidimensionais, considerando que os testes são multidimensionais. Ademais, coloca-se em discussão os avanços e possibilidades por esses modelos e suas implicações político-pedagógicas nas tomadas de decisão nas diferentes esferas administrativas dos governos.

Palavras-chave: Avaliações padronizadas; Teoria de resposta ao item; Política educacional.

ABSTRACT

This essay aimed to problematize possible political, pedagogical and methodological implications of large-scale educational tests. Therefore, we produced a survey of national and state educational tests administered nowadays in Brazil. The findings indicate the educational tests started to be implemented in the beginning of the 1990's, with classical psychometric being utilized at first. Along the years, governments began to deploy the unidimensional models of item response theory in result analysis as well as implemented tests more recently. These result analysis models of tests have been widely used in Brazil. Nevertheless, the authors question whether the tests adequation are related to the assumptions of the statistical model of unidimensional item response theory, given the fact that the tests are multidimensional. Furthermore, this study discusses the advances and possibilities by these models and their political and pedagogical implications in the decision making in the different administrative spheres of government.

Key-words: Standardized assessment; Item Response Theory; Educational Policy.

¹ Doutor em educação pela Universidade Federal do Ceará (UFC). Professor do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). ORCID: <https://orcid.org/0000-0002-0482-2699>. Email: leandro.sousa@ifce.edu.br

² Doutora em Educação pela Universidade Federal do Ceará (UFC). Professora da Universidade Federal do Ceará. ORCID: <https://orcid.org/0000-0001-5163-209X>. Email: adrianaufc@yahoo.com.br

1 INTRODUÇÃO

A avaliação de sistemas educacionais a partir de testes padronizados são importantes instrumentos de diagnóstico do aprendizado de estudantes em um contexto de larga escala (VIANNA, 2003). Com isso, esse instrumento tem sido amplamente utilizado no âmbito do governo federal e estadual brasileiro desde a década de 1990, proporcionando informações para as tomadas de decisão e implementação de políticas públicas. Mais recentemente, os governos municipais também têm cada vez mais aderidos a esses testes (BAUER et al., 2015).

Em suas primeiras aplicações, a administração desses instrumentos utilizava a teoria clássica dos testes (TCT) para analisar os resultados. No entanto, alguns problemas desse método são apontados, tais como a dependência que os parâmetros dos itens (dificuldade e discriminação) têm dos participantes que realizam o teste (ANDRADE; TAVARES; VALLE, 2000; ANDRADE; LAROS; GOUVEIA, 2010; MUÑIZ, 2010; SARTES; SOUSA-FORMIGONI, 2013). Segundo alguns desses autores essa instabilidade prejudica a comparação dos resultados ao longo dos anos, dificultando o acompanhamento do aprendizado dos estudantes.

Como forma de evitar esse problema passa a ser utilizado a teoria de resposta ao item (TRI) para a análise dos resultados desses testes (EMBRETSON; REISE, 2000; KLEIN, 2013; VALLE, 2000). Com a TRI é possível estimar parâmetros para os itens que independem dos participantes, colocando-os em uma mesma escala. Com isso, é possível estabelecer comparações dos participantes em aplicações diferentes e acompanhá-los ao longo do tempo. No entanto, algumas críticas são apresentadas a esse modelo, como o problema de adequação ao pressuposto da unidimensionalidade (TAVARES, 2013).

Também é importante destacar que esses testes podem induzir a elaboração de currículos escolares (SOUSA, 2003), além de influenciar a gestão (BERTAGNA; POLATO; MELLO, 2018) e a organização do trabalho docente e escolar (ALMEIDA; NHOQUE, 2017; AMARO, 2017). Em nível de sistema educacional, gestores também tem feito usos dos resultados das avaliações para implementações de políticas educacionais (SOUSA; OLIVEIRA, 2010). Muitas vezes essas ações ocorrem com finalidade de obtenção de bons resultados no exame.

Por esses testes influenciarem fortemente a rotina escolar, ou seja, a gestão da instituição, o trabalho pedagógico dos professores e, em um outro nível a gestão do sistema educacional nas diferentes esferas administrativas governamentais, consideramos

importante que medidas confiáveis sejam realizadas para que possam orientar adequadamente essas ações.

Nessa perspectiva, este ensaio objetivou problematizar as possíveis implicações políticas, pedagógicas e metodológicas dos testes educacionais em larga escala, principalmente como o método utilizado na construção e análise dos resultados dos testes podem gerar informações inadequadas e, conseqüentemente, tomadas de decisão desajustadas a realidade educacional.

2 TESTES PADRONIZADOS E AS AVALIAÇÕES DO SISTEMA EDUCACIONAL

As avaliações educacionais em grandes grupos populacionais surgem da necessidade de realização de prestação de contas de grandes investimentos financeiros aplicados nessa área, denominado de *accountability*, por vezes criticado como instrumento de regulação do Estado. Além disso, também da importância de obter diagnósticos que permitissem o planejamento e direcionamento de políticas focadas em tomadas de decisões para a resolução de problemas específicos no campo educacional.

Para tanto, é necessário remeter a ideia de qualidade. E quando se trata de educação, o conceito de qualidade não é fixo, depende dos sujeitos envolvidos e de suas necessidades (HORTA NETO, 2010). Dessa forma, o que é qualidade para um determinado grupo pode não o ser para outro. Portanto, ao realizar uma avaliação é necessário estabelecer os critérios de qualidade, ou seja, o padrão de referência para comparação dos seus resultados. Embora haja críticas às avaliações em larga escala por forçarem um padrão único de “qualidade” para o sistema educacional ao fazer uso apenas dos resultados das provas aplicadas (SOUSA, 2014). Quando falamos em qualidade do sistema, consideramos que este não é possível ser conhecido através de uma prova padronizada (CASASSUS, 2013), mas não podemos descartar que os indicadores revelados são aspectos necessários à ideia de qualidade em educação (MACHADO; ALAVARSE, 2014).

No Brasil, as avaliações em larga escala foram implantadas em diferentes níveis administrativos a partir da década de 1990, na tentativa de solucionar problemas e melhorar a qualidade da educação ofertada, que nesse contexto estava relacionada a proporcionar aquisição de competências e habilidades, ou seja, elevar o desempenho dos estudantes (VIANNA, 2003). Entre elas o Sistema de Avaliação da Educação Básica (Saeb), Prova Brasil, Exame Nacional do Ensino Médio (Enem) e Sistema Nacional de

Avaliação da Educação Superior (Sinaes). Essas avaliações assumem duas funções distintas (MINHOTO, 2016):

1) levantar informações ou evidências necessárias à formulação de políticas educacionais, tendo como propósito ampliar e aprofundar o conhecimento sobre os sistemas de ensino para que as diferentes esferas de governo possam definir prioridades de intervenção; e 2) induzir mudanças ou consolidar reformas educacionais previamente estruturadas para os sistemas de ensino (p. 78).

No entanto, há uma contradição importante levantada por Vianna (2003 p. 44) que permanece atualmente: “são desenvolvidas competências e habilidades em nosso sistema educacional de uma forma sistemática, ou, explicitando, é o nosso ensino orientado para o desenvolvimento de competências?”. Esse questionamento aponta para uma possível fragilidade do sistema educacional e de suas avaliações, uma vez de sua incoerência entre o ensinado e o avaliado. Além disso, ressalta que avaliações de grande abrangência populacional compromete uma avaliação completa, ou seja, que contemple todas as dimensões dos avaliados, sendo limitada a uma métrica do que se supõe medir.

Como forma de reverter este aspecto de limitação, Dalben e Almeida (2015) propõe uma perspectiva multidimensional para as avaliações em larga escala, de modo que pudesse abranger os mais variados fatores possíveis associados à aprendizagem dos alunos. No entanto, os autores reconhecem a impossibilidade de se contemplar todo o currículo da escala, uma vez que este sempre é mais abrangente que uma matriz de referência por mais ampla que seja. Além disso, ressaltam que com uma avaliação multidimensional as escolas podem eleger suas prioridades e metas a serem alcançadas considerando os aspectos avaliados, com isso lateraliza-se o foco exclusivo no rendimento.

Nesse contexto, Bauer, Alavarse e Oliveira (2015) realizam uma sistematização do debate e ressaltam que não há consenso em relação às contribuições das medidas educacionais em larga escala, mas reconhecem sua importância. Diante disso, os autores apontam duas linhas de discussão: i) o papel das avaliações em larga escala nas reformas educacionais; ii) avaliações educacionais como instrumento de gestão e política educacional.

Na primeira linha, Bauer, Alavarse e Oliveira (2015) reconhecem o papel das avaliações quando centrada no produto e não no processo. Nessa perspectiva, há argumentos favoráveis, uma vez que as escolas e professores são responsabilizados, estimulando o comprometimento destes com a aprendizagem dos estudantes. Além disso, instala uma cultura de avaliação, dando transparência aos resultados e processos do serviço

público. Também possibilita aos gestores a comparação de resultados com outras escolas ou com diferentes níveis administrativos, o que possibilita tomada de decisões e direcionamento de políticas públicas. Em contraponto, há severas críticas ao papel político das avaliações em larga escala, principalmente quando têm grandes impactos (*high stakes testing*). Argumenta-se no sentido que essas avaliações são instrumentos para imposição de reformas educacionais, aprofundando ainda mais a desigualdade. Além disso, denunciam os interesses econômicos associados, tais como terceirização dos testes e venda de materiais pela iniciativa privada.

Na segunda linha encontra-se os usos e divulgação dos resultados da avaliação. Nessa perspectiva, considera-se a importância da divulgação dos resultados (BAUER; ALAVARSE; OLIVEIRA, 2015). Por exemplo, quando proporciona aos alunos e professores informações sobre a aprendizagem, possibilitando tomadas de decisões. Aos pais, quando proporciona informações de qualidade de ensino das escolas, auxiliando na escolha da instituição em que os filhos serão matriculados. No entanto, há críticas de como a divulgação é realizada. Entre as objeções está em relação aos ranqueamentos das instituições, que não apresentam informações completas sobre a mesma. Outra questão colocada é em relação aos impactos causados as escolas e professores, quando é utilizada políticas de bonificação. Em decorrência disto, os professores acabam por adaptar seus métodos de ensino e limitam os conteúdos ensinados para atender as expectativas das avaliações.

Considerando esse viés político das avaliações em larga escala, Bonamino e Sousa (2012) apresenta a avaliação da educação básica no Brasil em três gerações, em que ressaltam principalmente o grau dos impactos sobre as escolas. Na primeira geração, as autoras apontam não haver impacto sobre as escolas e professores, tal como o Sistema de Avaliação da Educação Básica (Saeb), em que sua estrutura de avaliação visa diagnósticos gerais da educação no país, não permitindo comparações entre as escolas, deste modo apenas auxilia no direcionamento de políticas públicas para melhoria dos resultados. Em uma avaliação de segunda geração, já há alguma influência sobre as escolas, como é o caso da Prova Brasil, em que, por ser censitária, permite a comparação entre escolas e turmas, exercendo assim pressão sobre os gestores e professores para alcanças metas. Por fim, em uma avaliação de terceira geração há grandes impactos sobre escolas e professores, principalmente com as políticas de bonificação realizadas a partir dos resultados das avaliações, como é o caso dos sistemas de avaliação de São Paulo e Pernambuco. Como

consequência, pode haver alterações no processo de ensino e aprendizagem para melhorar os resultados.

3 ASPECTOS METODOLÓGICOS E IMPLICAÇÕES PEDAGÓGICAS DAS AVALIAÇÕES EM LARGA ESCALA.

Até aqui apresentamos o quanto que as avaliações em larga escala influenciam, em certa medida, as tomadas de decisão desde a gestão do sistema educacional por parte dos agentes governamentais e como essas avaliações, principalmente as de terceira geração (BONAMINO; SOUSA, 2012) podem influenciar a escola e seu agentes.

Considerando isto, para atender a rigorosidade necessária nos processos de construção e análise dos resultados, muitos exames e sistemas de avaliação no Brasil optaram por adotar a TRI como modelo de análise em detrimento da TCT, por considerarem permitir maior precisão e comparabilidade dos resultados. Diante disso, apresentamos o método (TCT e/ou TRI) de cálculo das proficiências (nota) dos participantes das avaliações estaduais e nacionais no Brasil (Quadro 1).

Quadro 1. Métodos de análise dos resultados nas avaliações nacionais e estaduais no Brasil.

Avaliações	Ano de Criação	Entes da Federação	Análise dos Resultados	
			Na criação	Atualmente
Avaliação Nacional da Educação Básica (Aneb)	1990	União	TCT	TRI
Avaliação Nacional do Rendimento Escolar (Aresc - Prova Brasil)	2005	União	TCT	TRI
Avaliação Nacional da Alfabetização (ANA)	2013	União	TRI	TRI
Exame Nacional do Ensino Médio (Enem)	1998	União	TCT	TRI
Sistema Estadual de Avaliação da Aprendizagem Escolar (Seape)	2009	Acre	TCT e TRI	TCT e TRI
Avaliação de Aprendizagem da Rede Estadual de Ensino de Alagoas (Areal)	2012	Alagoas	TCT e TRI	TCT e TRI
Sistema de Avaliação do Desempenho Educacional do Amazonas (Sadeam)	2008	Amazonas	TCT e TRI	TCT e TRI
Sistema de Avaliação Baiano da Educação (Sabe)	2011	Bahia	TCT e TRI	TCT e TRI
Sistema Permanente de Avaliação da Educação Básica do Ceará (Spaace)	1992	Ceará	TCT	TCT e TRI
Sistema Permanente de Avaliação Educacional do Distrito Federal (SipaeDF)	2014	Distrito Federal	TCT	TCT

Programa de Avaliação da Educação Básica do Espírito Santo (Paebes)	2009	Espírito Santo	TCT e TRI	TCT e TRI
Sistema de Avaliação Educacional do Estado de Goiás (Saego)	2011	Goiás	TRI	TRI
Sistema de Avaliação da Educação da Rede Pública de Mato Grosso do Sul (Saems)	2008	Mato Grosso do Sul	TCT e TRI	TCT e TRI
Avaliação Diagnóstica do Ensino Público do Estado de Mato Grosso (ADEPE-MT)	2016	Mato Grosso	TRI	TRI
Sistema Mineiro de Avaliação e Equidade da Educação Pública (SIMAVE)	2000	Minas Gerais	TCT	TCT e TRI
Sistema Paraense de Avaliação Educacional (SisPAE)	2013	Pará	TCT e TRI	TCT e TRI
Sistema Estadual de Avaliação da Educação da Paraíba (Avaliando IDEPB)	2012	Paraíba	TCT e TRI	TCT e TRI
Sistema de Avaliação da Educação Básica do Paraná (Saep)	2012	Paraná	TCT e TRI	TCT e TRI
Sistema de Avaliação Educacional de Pernambuco (Saepe)	2000	Pernambuco	TCT	TCT e TRI
Sistema de Avaliação Educacional do Piauí (Saepi)	2011	Piauí	TCT e TRI	TCT e TRI
Sistema de Avaliação da Educação do Estado do Rio de Janeiro (Saerj)	2008	Rio de Janeiro	TRI	TRI
Sistema Integrado de Monitoramento e Avaliação Institucional (Simais)	2016	Rio Grande do Norte	TCT e TRI	TCT e TRI
Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul (Saers)	1996	Rio Grande do Sul	TCT	TRI
Sistema de Avaliação Educacional de Rondônia (Saero)	2012	Rondônia	TCT e TRI	TCT e TRI
Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Saresp)	1996	São Paulo	TCT	TCT e TRI
Sistema de Avaliação da Educação do Estado do Tocantins (Saeto)	2011	Tocantins	TCT	TCT

Fonte: Elaboração própria.

Ao investigar os métodos de análise utilizados nas avaliações em larga escala no Brasil, atualmente, a grande maioria dos testes adotam a TRI na análise dos resultados, ou seja, no cálculo da proficiência dos participantes, embora em muitas dessas avaliações a TCT não tenha sido deixada totalmente de lado. Grande parte dessas avaliações podem ser consideradas como de terceira geração, ou seja, têm muita influência ou impacto sobre os interessados, a saber alunos, professores, gestão e a própria escola, considerando que muitas administrações públicas utilizam os resultados como política de bonificação dos mesmos, principalmente.

Nesse sentido, é importante que esses instrumentos apresentem informações confiáveis sobre o desempenho dos estudantes. Para tanto, é necessário dispor de métodos de análise dos resultados consistentes e que produzam boas informações. Diante disso, utilizar a TCT ou TRI pode influenciar nos resultados e em sua interpretação. Dessa forma, são importantes as evidências empíricas sobre a qualidade ou não desses métodos de análise dos resultados, principalmente da TRI, que tem sido cada vez mais adotada nos exames e avaliações educacionais.

Em relação a esse aspecto, além de algumas críticas ao modelo de TRI (TAVARES, 2013) mais utilizado atualmente já apresentado nesse texto, alguns estudos indicam um teste educacional amplamente utilizado no Brasil, como o Enem, que é parâmetro para seleção de candidatos aos cursos de graduação em todo o Brasil, não tem apresentado adequação ao seu pressuposto principal, o da unidimensionalidade (PRIMI; CICHETTO, 2018; SOUSA, 2019).

Este problema de adequação do modelo pode gerar interpretação errônea dos resultados e, conseqüentemente, prejudicar, no caso deste exame em específico, milhões de candidatos que realizam a prova anualmente. Em um esforço de ampliação da discussão, se essas inadequações ocorrerem em outras avaliações, com foco no sistema, como o Saeb, o modelo pode gerar interpretações equivocadas nos resultados. Desse modo, os gestores políticos e educacionais, assim como professores estariam sendo orientados e, por consequência, tomando decisões de forma equivocada.

4 À GUIA DE CONCLUSÃO

As avaliações educacionais ganharam relativa importância no cenário educacional nos últimos anos. Tem contribuído significativamente para o acompanhamento dos sistemas educacionais em todos os âmbitos, municipal, estadual e federal, oferecendo informações que possibilitam indicadores para as tomadas de decisão dos gestores políticos e educacionais, orientando currículos e estimulando inovações didáticas nas escolas. Ademais, é necessário a compreensão de que as avaliações, principalmente as realizadas em larga escala, são imbuídas de aspectos técnicos e políticos que, por vezes, influenciam a atividade pedagógica da escola e de seus atores.

Os aspectos técnicos das avaliações em larga escala, especificamente os modelos estatísticos empregados na análise e correção da prova devem ser avaliados cuidadosamente, principalmente em relação à validade e utilidade dos resultados por eles

gerados. Do contrário essas avaliações podem não atingir o propósito para o qual foram elaboradas.

6 REFERÊNCIAS

ALMEIDA, J. G.; NHOQUE, J. R. Avaliação em larga escala na escola: repercussões e desdobramentos no trabalho escolar Júlio. **Revista de Estudos Aplicados em Educação**, v. 2, n. 4, p. 84-95, 2017. Doi: <https://doi.org/10.13037/rea-e.vol2n4.4781>.

AMARO, I. Avaliação em larga escala e trabalho docente: da lógica efficientista à lógica contrarregulatória. **Quaestio: revista de estudos em educação**, v. 19, n. 2, p. 417, 4 set. 2017. Doi: <https://doi.org/10.21723/riaee.v11.n4.7155>.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. DA C. **Teoria de Resposta ao Item: Conceitos e Aplicações**. [s.l.] ABE – Associação Brasileira de Estatística, 2000.

ANDRADE, J. M.; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. **Avaliação Psicológica**, v. 9, n. 3, p. 421-435, 2010. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300009>. Acesso em: 29 maio 2020.

BAUER, A.; PIMENTA, C. O.; HORTA NETO, J. L.; SOUSA, S. Z. L. Avaliação em larga escala em municípios brasileiros: o que dizem os números? **Estudos em Avaliação Educacional**, v. 26, n. 62, p. 326-352, 2015. Doi: <http://dx.doi.org/10.18222/eae266203207>.

BAUER, A.; ALAVARSE, O. M.; OLIVEIRA, R. P. Avaliações em larga escala: uma sistematização do debate. **Educ. Pesqui.**, v. 41, n. especial, p. 1367-1382, 2015. Doi: <https://doi.org/10.1590/S1517-9702201508144607>.

BERTAGNA, R. H.; POLATO, A.; DE MELLO, L. R. As avaliações em larga escala e o currículo no estado de São Paulo: A realidade escolar desvelando as lógicas subjacentes. **Currículo Sem Fronteiras**, v. 18, n. 1, p. 52-70, 2018. Disponível em: <<http://www.curriculosemfronteiras.org/vol18iss1articles/bertagna-polato-mello.pdf>>. Acesso em: 29 maio. 2020.

BONAMINO, A.; SOUSA, S. Z. L. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. **Educação e Pesquisa**, v. 38, n. 2, p. 373-388, 2012. Doi: <https://doi.org/10.1590/S1517-97022012005000006>.

CASASSUS, J. Política y metáforas: un análisis de la evaluación estandarizada en el contexto de la política educativa. **Vinte e cinco de avaliação de sistemas educacionais no Brasil: origens e pressupostos. Florianópolis: Insular**, p. 21-46, 2013.

DALBEN, A.; ALMEIDA, L. C. Para uma avaliação de larga escala multidimensional. **Est. Avali. Educ.**, v. 26, n. 61, p. 12-28, 2015. Doi: <http://dx.doi.org/10.18222/eae266103140>.

EMBRETSON, S. E.; REISE, S. P. **Item response theory for psychologists**. Mahwah, N.J: Lawrence Erlbaum Associates, Inc., 2000.

HORTA NETO, J. L. Avaliação externa de escolas e sistemas: questões presentes no debate sobre o tema. **R. bras. Est. pedag.**, v. 91, n. 227, p. 84-104, 2010. Doi: <http://dx.doi.org/10.24109/2176-6681.rbep.91i227.604>.

KLEIN, R. Alguns aspectos da Teoria de Resposta ao Item relativos à estimação das proficiências. **Ensaio: Aval. Pol. Públ. Educ.**, v. 21, n. 78, p. 35-56, 2013. Doi: <https://doi.org/10.1590/S0104-40362013005000003>.

MACHADO, C.; ALAVARSE, O. M. Qualidade das Escolas: tensões e potencialidades das avaliações externas. **Educação & Realidade**, v. 39, n. 2, p. 413-436, 2014. Doi: <https://doi.org/10.1590/S2175-62362014000200005>.

MINHOTO, M. A. P. Política de Avaliação da Educação Brasileira : limites e perspectivas. **Jornal de Políticas Educacionais**, v. 10, n. 19, p. 77-90, 2016. Doi: <http://dx.doi.org/10.5380/jpe.v10i19.50214>.

MUÑIZ, J. Las teorías de los tests: Teoría clásica y teoría de respuesta a los ítems. **Papeles del Psicólogo**, v. 31, n. 1, p. 57-66, 2010. Disponível em: <<http://www.papelesdelpsicologo.es/pdf/1796.pdf>>. Acesso em 29: maio 2020.

PRIMI, R.; CICHETTO, A. A. **Como os escores do ENEM são atribuídos pela TRI? Anais do VI CONBRATRI: Métodos para detecção de fraudes em testes** Juiz de Fora, 2018. Disponível em: <<https://even3.blob.core.windows.net/anais/92067.pdf>>. Acesso em: 29 maio 2020.

SARTES, L. M. A.; SOUSA-FORMIGONI, M. L. O. Avanços na Psicometria: da teoria clássica dos testes à teoria de resposta ao item. **Psicologia: Reflexão e Crítica**, v. 26, n. 2, p. 241-250, 2013. Doi: <https://doi.org/10.1590/S0102-79722013000200004>.

SOUSA, L. A. **Análise comparativa do exame nacional do ensino médio (Enem) via teoria clássica do teste e teoria de resposta ao item**. Tese (Doutorado) - Universidade Federal do Ceará, Programa de Pós-graduação em Educação, Fortaleza(CE), 2019. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/48275>>. Acesso em: 29 maio 2020

SOUSA, S. Z. L. Possíveis impactos das políticas de avaliação no currículo escolar. **Cadernos de Pesquisa**, n. 119, p. 175-190, 2003. Doi: <https://doi.org/10.1590/S0100-15742003000200009>.

SOUSA, S. Z. L. Concepções de qualidade da educação básica forjadas por meio de avaliações em larga escala. **Avaliação**, v. 14, n. 2, p. 407-420, 2014. Doi: <https://doi.org/10.1590/S1414-40772014000200008>.

SOUSA, S. Z. L.; OLIVEIRA, R. P. Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências. **Cadernos de Pesquisa**, v. 40, n. 141, p. 793-822, dez. 2010. Doi: <https://doi.org/10.21573/vol31n32015.6380>.

TAVARES, C. Z. Teoria da resposta ao item: uma análise crítica dos pressupostos epistemológicos. **Estudos em Avaliação Educacional**, v. 24, n. 54, p. 56-76, 2013. Doi: <http://dx.doi.org/10.18222/eae245420131902>.

VALLE, R. C. Teoria de resposta ao item. **Estudos em Avaliação educacional**, v. 21, p. 07-92, 2000. Doi: <http://dx.doi.org/10.18222/eae02120002225>.

VIANNA, H. M. Avaliações nacionais em larga escala: análises e propostas. **Estudos em Avaliação Educacional**, n. 27, p. 41-76, 2003. Doi: <http://dx.doi.org/10.18222/eae02720032177>.