

ANÁLISE DA CONFIABILIDADE DA ESCALA DE AVALIAÇÃO DOS AMBIENTES DE APRENDIZAGEM DEDICADOS À PRIMEIRA INFÂNCIA (EAPI) E O IMPACTO DAS CARACTERÍSTICAS DOS APLICADORES

José Carlos Souza Lima¹
Doutorando em Caen/UFC
Fortaleza, Ceará, Brasil
jose.carlos@caen.ufc.br

Guilherme Irffi²
Professor do Departamento de
Economia Aplicada e do Caen/UFC
Fortaleza, Ceará, Brasil
irffi@caen.ufc.br

Resumo

A qualidade na educação infantil pode ser refletida nas interações responsivas entre adultos e crianças e nos ambientes de aprendizagem. Para aprimorá-la, é essencial o uso de instrumentos de monitoramento. Este artigo analisa a confiabilidade da Escala de Avaliação dos Ambientes de Aprendizagem da Primeira Infância (EAPI) em formato contínuo, considerando equivalência, estabilidade e consistência interna. Também são examinadas características dos aplicadores como variáveis explicativas dos indicadores de confiabilidade. Os resultados mostram que o número de aplicações não afeta a equivalência nem a estabilidade, mas o número de aplicadores influencia a confiabilidade. Reaplicações com dois ou três aplicadores tendem a gerar melhores coeficientes. A consistência interna é impactada negativamente por aplicações contínuas, mas tende a se estabilizar após a quarta aplicação. Por fim, conhecimento sobre desenvolvimento infantil, experiência e bom desempenho na certificação contribuem positivamente para a confiabilidade, enquanto maior carga de aplicações semanais exerce efeito negativo.

Palavras chaves: Educação infantil; Confiabilidade; Medidas de Qualidade; Ambientes de aprendizagem.

1 INTRODUÇÃO

Investir na primeira infância é garantir um direito fundamental previsto na Constituição Federal (Art. 227) e no Estatuto da Criança e do Adolescente (Art. 4), que asseguram prioridade absoluta a esse grupo formulação e execução de políticas sociais públicas. Além disso, os retornos sociais e econômicos desse investimento podem superar os de intervenções em fases posteriores (Carneiro; Heckman, 2003).

Crianças que frequentam a Educação Infantil (EI) tendem a apresentar melhores resultados em alfabetização, habilidades sociais, desempenho escolar e saúde, além de menores taxas de evasão e criminalidade (Shirasu *et al.*, 2022; Camilli *et al.*, 2010; Irffi; Pereira, 2022; Reynolds *et al.*, 2002; Campbell *et al.*, 2014).

Com a ampliação do acesso à EI, a atenção volta-se à qualidade da oferta, que apesar de não ser uma tarefa simples, pode ser entendida como a garantia de um ambiente seguro,

¹ O autor agradece a CAPES pelo financiamento.

² O autor agradece ao CNPQ pelo financiamento.

acolhedor e estimulante ao desenvolvimento integral (Yoshikawa *et al.*, 2018; Castilho; Ogando; Gil, 2021).

Assim, para aumentar a eficácia da EI, é necessária uma ampla gama de iniciativas que alie a promoção de espaços físicos e sociais que ofereçam condições propícias para o desenvolvimento integral das crianças e apoios direcionados às interações dos professores com as crianças (Early *et al.*, 2007). Uma maneira de elevar a qualidade é através da realização regular de monitoramentos em todos os contextos da primeira infância, com o intuito de identificar e registrar práticas de alto padrão (Yoshikawa *et al.*, 2018).

Melhorar a qualidade requer ambientes adequados e interações qualificadas, o que pode ser promovido por meio do monitoramento contínuo (Early *et al.*, 2007; Yoshikawa *et al.*, 2018). Nesse contexto, instrumentos de avaliação confiáveis e válidos são essenciais. Porém, sua eficácia também depende das condições e capacidades dos aplicadores (Shirasu *et al.*, 2021; Vitiello *et al.*, 2018).

A Escala de Avaliação dos Ambientes de Aprendizagens dedicados à Primeira Infância (EAPI)³ busca mensurar aspectos estruturais e processuais da qualidade na EI. Ela pode ser dividida em dimensões: Alimentação; Currículo, interações e práticas; Diversidade funcional; Infraestrutura; e Segurança. Essas dimensões são compostas por itens que representam práticas e condições observáveis no cotidiano das instituições de educação infantil.

Por exemplo, a dimensão de 'Currículo, interações e práticas' avalia a presença de atividades baseadas no brincar e nas interações entre adultos e crianças, enquanto 'Infraestrutura' observa elementos físicos do ambiente, como ventilação e mobiliário adequado. A concepção dessas escalas baseia-se em literatura nacional e internacional sobre qualidade em contextos de primeira infância.

O objetivo deste trabalho é analisar a confiabilidade da aplicação do instrumento de observação da EAPI por meio da equivalência, estabilidade e consistência interna (Souza *et al.*, 2017; Pilatti *et al.*, 2010; Downer *et al.*, 2010). Além de investigar sobre os aplicadores e a quantidade de reaplicações utilizando características individuais e informações da coleta de campo como variáveis explicativas para os indicadores da confiabilidade (Pianta; La Paro; Hamre, 2008; Downer *et al.*, 2010).

A análise de confiabilidade fornece evidências importantes sobre o desempenho do instrumento, reconhece-se que ela faz parte do processo de validação. Este estudo, no entanto,

³ Disponível em: <https://www.observatorioei.org.br/>

concentra-se especificamente na confiabilidade da EAPI, em razão da riqueza dos dados disponíveis no formato contínuo de aplicação, que permite observar variações decorrentes de reaplicações, da distribuição dos aplicadores entre as turmas e das próprias características individuais desses profissionais. Esses elementos fornecem possibilidades para investigar aspectos como experiência, conhecimento prévio e carga de trabalho dos aplicadores influenciam a consistência das avaliações realizadas por meio da EAPI.

2 METODOLOGIA

Serão utilizados os dados referentes a aplicação denominada EAPI contínua, no qual o instrumento de observação foi aplicado cinco vezes na mesma turma, em um intervalo de tempo menor que um mês. Os itens da EAPI foram agregados em dimensões para a análise dos resultados, são elas: Alimentação; Currículo, interações e práticas; Diversidade funcional; Infraestrutura; e Segurança.

A confiabilidade da EAPI, dado esse formato, se refere à capacidade do instrumento produzir resultados consistentes ao longo do tempo, entre diferentes observadores e em diferentes contextos. Existem três critérios principais de confiabilidade: equivalência, estabilidade e consistência interna (Souza *et al.*, 2017). Cada critério possui métodos estatísticos específicos para sua avaliação.

A equivalência refere-se à concordância entre observadores sobre os escores de um instrumento, para essa medida foi usado o coeficiente Kappa. A estabilidade refere-se à consistência dos resultados obtidos em diferentes momentos no tempo e foi estimada usando o coeficiente de correlação intraclass (ICC). E a consistência interna indica se todas as partes de um instrumento medem a mesma característica, coeficiente alfa de Cronbach é comumente utilizado para essa análise.

A estimação do modelo com as características individuais dos observadores foi realizada utilizando o seguinte modelo MQO com efeitos fixos:

$$Y_{it} = \alpha + \beta X_{it} + \mu_i + \delta_t + \varepsilon \quad (1)$$

Onde Y_{it} são os coeficientes de confiabilidade, X_{it} as variáveis explicativas. μ_i e δ_t são os efeitos fixos individuais e temporais, respectivamente. ε é o termo de erro.

3 ANÁLISE E DISCUSSÃO DOS RESULTADOS

À medida que a quantidade de aplicadores aumenta, diminui a concordância (Kappa) entre eles. Isso é um movimento esperado e pode ser ocasionado pelo efeito do acaso ou por discrepâncias nas avaliações (Fleiss, Nee e Landis, 1979). Quando um mesmo indivíduo

realiza todas as aplicações em uma turma, ele pode se habituar ao ambiente e aprimorar suas respostas sobre os insumos. No entanto, também pode recorrer à memória, em vez da observação direta, ao preencher alguns itens.

Quanto aos níveis de concordância, no caso em que é o mesmo observador em todas as aplicações na turma, as dimensões apresentam resultados *excelentes*⁴, exceto a de “Currículos, interações e práticas” que foi nível *bom* de concordância. Quando são dois observadores, todas as dimensões se deslocam um nível abaixo. Adicionar uma visita a mais quando se utiliza mais de dois observadores para aplicar a EAPI na turma pode reduzir marginalmente a concordância em um nível.

O ICC foi estimado de forma semelhante ao Kappa e os resultados apresentam um comportamento semelhante. Porém, a dimensão “Currículo, interações e práticas” permanece com nível de estabilidade, considerado *bom*, mesmo com mais aplicações e observadores. Já “Segurança” chegou a alcançar um nível *excelente* para um observador e *bom* para dois ou mais. As outras três dimensões apresentaram nível *bom* de estabilidade “intraobservador” e *moderado* para dois ou mais aplicadores. Isso sugere que reaplicações contínuas da EAPI em uma turma possuem correlações aceitáveis, quando o intervalo de tempo for de até 31 dias⁵.

Foi realizado a estimação do alfa para cada turma. As dimensões de “Currículo, interações e práticas” e “Infraestrutura” apresentaram valores relativamente altos para o alfa, com média de 0,93 e 0,90, respectivamente, nível considerado *perfeito*. As demais dimensões apresentam, em média, *boa* consistência interna, com pelo menos 70% das turmas nesse nível ou superior e menos de 15% nos níveis *fraco* ou *insignificante*. Esses resultados se aproximam do que é mostrado por Dockrell *et al.* (2023) para o MELE⁶. A consistência interna melhora quanto mais diversificados forem os observadores em reaplicações e em geral os constructos medidos pela EAPI demonstraram consistência interna *boa* ou *excelente*.

De acordo com a estimação de (1), a concordância “interobservador” sofreu um impacto estatisticamente significativo da quantidade de observadores, da distância, do conhecimento sobre desenvolvimento infantil, da experiência do aplicador, da nota da certificação da EAPI e do número de observações por semana. O aumento no número de observadores implica em uma redução marginal no valor de Kappa de 0,03; 0,05 e 0,14 nas dimensões de “Currículo, interações práticas”, “Diversidade funcional” e “Segurança”,

⁴ Interpretação dos coeficientes adaptada de Freitas & Rodrigues (2005), Cicchetti (1994) e Landis & Koch (1977).

⁵ Intervalo máximo de tempo entre a primeira e quinta aplicação da EAPI.

⁶ Instrumento no qual a EAPI foi adaptada.

respectivamente. A distância percorrida não teve impacto, por outro lado, a experiência apresentou efeito de 0,08 na dimensão de “Segurança”.

O número de observações impactou de forma negativa em aproximadamente 0,02 pontos marginais do ICC. É um valor consideravelmente baixo, o que torna possível explorar cenários contínuos de aplicação do instrumento, desde que haja clareza nos objetivos. As demais variáveis não apresentaram valores estatisticamente significantes.

Para concluir, o coeficiente Alfa de Cronbach do instrumento sofreu efeito, estatisticamente significativo, da quantidade de observações em todas as dimensões. Em suma, uma aplicação a mais tende a reduzir a consistência interna da coleta em até 0,06 pontos. “Currículo, interações e práticas” foi a dimensão menos afetada, com queda de 0,02 aproximadamente. Isso significa que se um dos focos do uso contínuo for a ampliação da observação de oportunidades de aprendizagem, é possível realizar reaplicações com uma margem baixa de redução da consistência do instrumento.

A experiência e a nota de certificação impactaram de forma negativa no Alfa de Cronbach para a dimensão de “Segurança”. Isso sugere que os itens que compõem esse constructo reduzem a capacidade de medi-lo dado o desempenho do observador na certificação e sua experiência na coleta dos dados. Dado que a EAPI se propõe a ter uma capacitação padronizada, abrem-se pretextos para um estudo mais aprofundado desse processo.

4 CONCLUSÃO

Este artigo analisou a confiabilidade da EAPI, avaliando sua estabilidade, equivalência e consistência interna. Também foram consideradas características dos observadores — como conhecimento sobre desenvolvimento infantil, desempenho na certificação, experiência e volume semanal de observações — na estimativa dos coeficientes de confiança.

Os resultados mostraram que o formato contínuo manteve a confiabilidade em grande parte dos cenários. O número de reaplicações não afetou a concordância entre observadores nem a estabilidade, mas o número de aplicadores influenciou todas as medidas de confiabilidade. A consistência interna foi impactada negativamente por aplicações contínuas, mas estabilizou-se após o quarto dia de observação.

Em reaplicações, dois ou três observadores apresentaram melhores resultados, embora mais observações tendam a reduzir os coeficientes. Recomenda-se um formato com duas a quatro aplicações, ou ainda a combinação de estratégias diferenciadas como um grupo de

turma receber mais observações, outro receber menos, um receber dois observadores simultaneamente e outro receber apenas um por vez.

5 REFERÊNCIAS

- CAMILI, G *et al.* Meta-analysis of the effects of early education interventions on cognitive and social development. **Teachers College Record**, v. 112, n. 3, p. 579-620, 2010.
- CAMPBELL, F *et al.* Early childhood investments substantially boost adult health. **Science**, v. 343, n. 6178, p. 1478-1485, 2014.
- CARNEIRO, P., HECKMAN, J. J. Human capital policy. In J. J. Heckman, A. B. Krueger, & B. M. Friedman (Eds.), *Inequality in America: What role for human capital policies?* p. 77-239. **MIT Press**, 2003.
- CASTILHO, Priscila Carvalho de; OGANDO, Laura Duarte; GIL, Marcia de Oliveira Gomes. **Educação Infantil de Qualidade**. Comitê Científico do Núcleo Ciência Pela Infância. São Paulo: Fundação Maria Cecília Souto Vidigal, 2021.
- DOCKRELL, Julie *et al.* Measurement of cognition and profiling early learning environments in India, Indonesia and Senegal: a UKRI GCRF Action Against Stunting Hub protocol paper. **BMJ Paediatrics Open**, 2023.
- EARLY, Diane M. *et al.* Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. **Child development**, v. 78, n. 2, p. 558-580, 2007.
- FLEISS, Joseph L.; NEE, John C.; LANDIS, J. Richard. Large sample variance of kappa in the case of different sets of raters. **Psychological bulletin**, v. 86, n. 5, p. 974, 1979.
- IRFFI, G.; PEREIRA, M. A. F. O impacto da Educação Infantil sobre o desempenho dos alunos brasileiros no Pisa. **Anais do 50º Encontro Nacional de Economia - ANPEC**, Fortaleza/CE, 2022.
- PIANTA, Robert C.; LA PARO, Karen M.; HAMRE, Bridget K. **Classroom Assessment Scoring System™: Manual K-3**. Paul H Brookes Publishing, 2008.
- PILATTI, L. A.; PEDROSO, B.; GUTIERREZ, G, L. Propriedades Psicométricas de Instrumentos de Avaliação: Um debate necessário. **Revista Brasileira de Ensino de Ciência e Tecnologia**, v. 3, n. 1, 2010. Disponível em: <<https://periodicos.utfpr.edu.br/rbect/article/view/619>>
- REYNOLDS, A. J. *et al.* Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers. **Educational Evaluation and Policy Analysis**, v. 24 n. 4, p. 267-303, 2002.
- SOUZA, A. C; ALEXANDRE, N.; GUIRARDELLO, E. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. **Epidemiol. Serv. Saúde, Brasília**, v. 26, n. 3, p. 649-659, set. 2017.
- SHIRASU, M. *et al.* Análise do impacto da educação infantil sobre a alfabetização nos municípios brasileiros, e seus efeitos heterogêneos nos municípios sob área de atuação da SUDENE. **Anais do XXVII Encontro Regional de Economia**, Fortaleza/CE, 2022.
- SHIRASU, M. *et al.* Medidas de Qualidade da Educação Infantil: Uma Revisão da Literatura, Universidade Federal do Ceará e Sudene, 2021. (Relatório Técnico, TED 42020).
- VITIELLO, Virginia E. *et al.* Measuring the quality of teacher-child interactions at scale: Comparing research-based and state observation approaches. **Early Childhood Research Quarterly**, v. 44, p. 161-169, 2018.
- YOSHIKAWA, Hirokazu *et al.* Toward high-quality early childhood development programs and policies at national scale: Directions for research in global contexts. **Social Policy Report**, v. 31, n. 1, p. 1-36, 2018.