



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA - CAEN**  
**MESTRADO ACADÊMICO EM ECONOMIA**

**JOSÉ CARLOS SOUZA LIMA**

**ENSAIOS SOBRE A CONFIABILIDADE DA ESCALA DE AVALIAÇÃO DOS  
AMBIENTES DE APRENDIZAGEM DEDICADOS À PRIMEIRA INFÂNCIA**

**FORTALEZA**

**2024**

JOSÉ CARLOS SOUZA LIMA

ENSAIOS SOBRE A CONFIABILIDADE DA ESACALA DE AVALIAÇÃO DOS AMBIENTES  
DE APRENDIZAGEM DEDICADOS À PRIMEIRA INFÂNCIA

Dissertação apresentada ao Programa de Pós-Graduação em Economia – CAEN da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Economia.

Orientador: Prof. Guilherme Irffi.

FORTALEZA

2024

ENSAIOS SOBRE A CONFIABILIDADE DA ESACALA DE AVALIAÇÃO DOS AMBIENTES  
DE APRENDIZAGEM DEDICADOS À PRIMEIRA INFÂNCIA

Dissertação apresentada ao Programa de Pós-Graduação em Economia – CAEN da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Economia

Orientador: Prof. Guilherme D. Irffi.

Aprovada em: 16/02/2024.

BANCA EXAMINADORA

---

Prof. Guilherme D. Irffi (Orientador)

Universidade Federal do Ceará (UFC)

---

Profª. Dra. Aline Maria Gomes Lima

Secretaria Municipal de Educação de Fortaleza (SME)

---

Prof. Dr. Daniel Domingues dos Santos

Universidade de São Paulo (USP-RP)

---

Prof. Dr. Diego Rafael Fonseca Carneiro

Universidade Federal do Ceará (UFC)

À minha mãe e ao meu avô, Rosilene e Estácio.

## AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Em primeiro lugar à minha mãe e ao meu avô, que agora nos olha do céu, Rosilene Costa e Estácio Costa que sempre estiveram ao meu lado e no meu coração em todos os momentos, mesmo à distância.

A minha namorada, Brysa Fernandes, que foi minha parceira nos bons e maus momentos desde o primeiro mês de aula da graduação.

Ao meu orientador, Guilherme Irffi, por todos os ensinamentos, oportunidades e por acreditar no meu trabalho.

Aos amigos que tive a oportunidade de conhecer e conviver durante essa jornada que é o mestrado acadêmico, Antônio Victor, Guilherme Côrrea, Letícia Lima, Pamella Nogueira, Yuri Timbó e Wendel.

Aos meus amigos de longa data que permanecem em minha vida, Cícero Lima e Ramon Rodrigues.

Aos meus amigos Analice Sampaio, Angelica Caitano, Isabela Braga, Marcos Renan, Mayana Andrade, Odalice Sampaio e Tony Araújo.

Ao Laboratório de Estudos e Pesquisas em Educação e Economia Social (LEPES) e ao Centro de Gestão Para Resultados (CGPR).

Aos meus gatos, Eva e Marvin, que estiveram ao meu lado durante toda a escrita desse trabalho, proporcionando uma excelente companhia.

À Instituição UFC, que me possibilitou cursar uma pós graduação.

Aos professores participantes da banca examinadora, Aline Maria, Daniel dos Santos e Diego Carneiro, pelo tempo e pelas valiosas colaborações e sugestões.

## RESUMO

A qualidade na educação infantil pode ser expressa pelas interações responsivas entre adultos e crianças e pelos ambientes de aprendizagem, que por sua vez devem ser seguros, estimulante e integral para proporcionar o desenvolvimento infantil (Raikes *et al.*, 2023; Yoshikawa *et al.*, 2013; Maia e Williams, 2005). Aprimorar essa qualidade é possível através do monitoramento (Yoshikawa *et al.*, 2018), e para isso, faz-se necessário o uso de instrumentos para a coleta de dados. Este trabalho é composto por dois ensaios, cujo os objetivos são explorar a confiabilidade da Escala de Avaliação dos Ambientes de Aprendizagens dedicados à Primeira Infância (EAPI), que busca mensurar a qualidade na educação infantil. O primeiro ensaio busca analisar a confiabilidade da aplicação da EAPI no formato contínuo de cinco dias de aplicação, por meio da estabilidade (ao longo do tempo), consistência interna (do instrumento) e equivalência (concordância dos observadores) (Souza *et al.*, 2017). O segundo utiliza um painel de dados para realizar uma análise das características dos observadores e o impacto delas na confiabilidade das respostas dos aplicadores ao instrumento. Foram utilizados os conhecimentos prévios dos observadores sobre desenvolvimento infantil e informações a respeito da coleta de campo. Foi encontrado que o número de aplicações não afeta a equivalência e a estabilidade, por outro lado, o número de observadores influencia a confiabilidade. Em reaplicações, o ideal é que seja diversificado o observador para dois ou três indivíduos diferentes, para melhores coeficientes. A consistência interna sofreu efeito negativo de aplicações contínuas, mas indica estabilizar-se depois de quatro dias. Por fim, o conhecimento sobre desenvolvimento infantil, experiência e o desempenho na certificação contribuíram positivamente para confiabilidade e a carga de aplicações semanais negativamente.

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

**Código JEL:** C23; I21; I24; J24.

## ABSTRACT

The quality of early childhood education can be expressed through responsive interactions between adults and children and by learning environments, which in turn must be safe, stimulating, and holistic to foster child development (Raikes *et al.*, 2023; Yoshikawa *et al.*, 2013; Maia & Williams, 2005). Enhancing this quality is possible through monitoring (Yoshikawa *et al.*, 2018), requiring the use of instruments for data collection. This work comprises two essays aimed at exploring the reliability of the “Escala de Avaliação dos Ambientes de Aprendizagens dedicados à Primeira Infância” (EAPI), which seeks to measure quality in early childhood education. The first essay seeks to analyze the reliability of the EAPI application in the continuous format of five days of application, through stability (over time), internal consistency (of the instrument), and equivalence (agreement among observers). The second uses a data panel to conduct an analysis of observer characteristics and their impact on the reliability of applicators' responses to the instrument. Previous knowledge of observers about child development and information regarding field data collection were used. It was found that the number of applications does not affect equivalence and stability; however, the number of observers influences reliability. In reapplications, it is ideal to diversify the observer to two or three different individuals, for better coefficients. Internal consistency was negatively affected by continuous applications but indicates stabilizing after four days. Finally, knowledge about child development, experience, and performance in certification contributed positively to reliability, while the weekly application load negatively impacted it.

**Keywords:** Early Childhood Education; Reliability; Quality Measures; Learning Environments.

**Jel Code:** C23; I21; I24; J24.

## **LISTA DE FIGURAS**

<b>Figura 1</b> - Municípios participantes do estudo nacional (2021). .....	18
<b>Figura 2</b> – Relações entre confiabilidade e validade. ....	25

## LISTA DE GRÁFICOS

<b>Gráfico 1</b> – Quantidades de observações realizadas em cada semana.....	50
<b>Gráfico 2</b> – Distribuição dos coeficientes Kappa das turmas por dimensão. ....	72
<b>Gráfico 3</b> – Distribuição dos coeficientes ICC das turmas por dimensão. ....	72
<b>Gráfico 4</b> – Distribuição dos coeficientes alfa de Cronbach das turmas por dimensão. ....	73
<b>Gráfico 5</b> – Densidade das distâncias percorridas pelos observadores de casa para a Unidade Educacional. ....	73

## LISTA DE TABELAS

<b>Tabela 1</b> – Estatísticas descritivas da base de dados. ....	23
<b>Tabela 2</b> – Coeficiente Kappa a nível de rede por dia observado e por faixa etária. ....	29
<b>Tabela 3</b> – Exemplo de matriz construída para cálculo do Coeficiente Kappa de uma turma. ....	31
<b>Tabela 4</b> – Coeficiente Kappa a nível de turma por dimensão e dia adicional de observação. ....	31
<b>Tabela 5</b> – Coeficiente Kappa a nível de turma por seções que compõe a dimensão currículo, interações e práticas. ....	32
<b>Tabela 6</b> – Coeficiente Kappa a nível de turma: análise das seções “Planejamento e currículo” e “Práticas de ampliação do repertório por meio do brincar”, ....	33
<b>Tabela 7</b> – Estimação do Kappa a nível de turma explorando o formato contínuo de aplicação. ....	34
<b>Tabela 8</b> – ICC estimado a nível de turma. ....	36
<b>Tabela 9</b> – Estatísticas descritivas das durações entre as aplicações da EAPI em cada turma. ....	37
<b>Tabela 10</b> – ICC estimado a nível de turma por quantidade de aplicações em cada intervalo de tempo. ....	38
<b>Tabela 11</b> – ICC estimado a nível de turma por quantidade de aplicações e de observadores. ....	39
<b>Tabela 12</b> – Alfa de Cronbach a nível de rede por aplicação e faixa etária. ....	40
<b>Tabela 13</b> – Alfa de Cronbach a nível de turma (n=144). ....	41
<b>Tabela 14</b> – Alfa estimado a nível de turma por quantidade de aplicações e de observadores. ....	42
<b>Tabela 15</b> – Alfa de Cronbach reestimado removendo um item. ....	43
<b>Tabela 16</b> – Estatísticas descritivas dos dados utilizados. ....	49
<b>Tabela 17</b> – Correlação entre as características dos observadores. ....	50
<b>Tabela 18</b> – Resultados da estimação para o coeficiente Kappa. ....	52
<b>Tabela 19</b> – Resultados da estimação para o ICC. ....	53
<b>Tabela 20</b> – Resultados da estimação para o Alfa de Cronbach. ....	54

## LISTA DE QUADROS

<b>Quadro 1</b> – Dimensões da EAPI. ....	19
<b>Quadro 2</b> – Interpretação par os coeficientes de confiabilidade. ....	28
<b>Quadro A2</b> – Itens da EAPI que compõem as dimensões, seções e facetas. ....	64

## LISTA DE ABREVIATURAS

CEI	Centro de Educação Infantil
EAPI	Escala de Avaliação dos Ambientes Dedicados à Primeira Infância
EI	Educação Infantil
EMTP	Escola Municipal de Tempo Parcial
MELE	Measure of Early Learning Environments
MELQO	Measuring Early Learning Quality and Outcomes
MODEL	Measure of Development of Early Learning
PARMA	Práticas de ampliação do repertório por meio do brincar
PC	Planejamento e Currículo
PNE	Plano Nacional de Educação
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children's Emergency Fund.

## SUMÁRIO

<b>1. INTRODUÇÃO GERAL</b> .....	12
<b>2. ESCALA DE AVALIAÇÃO DOS AMBIENTES DE APRENDIZAGENS DEDICADOS À PRIMEIRA INFÂNCIA (EAPI)</b> .....	15
<b>2.1 Measuring Early Learning Environment (MELE)</b> .....	15
<b>2.2 Elaboração da EAPI e a EAPI contínua</b> .....	17
<b>3. ENSAIO I: UMA ANÁLISE DA CONFIABILIDADE DA APLICAÇÃO CONTÍNUA DA ESCALA DE AVALIAÇÃO DOS AMBIENTE DE APRENDIZAGEM DEDICADOS À PRIMEIRA INFÂNCIA (EAPI)</b> .....	22
<b>3.1. Introdução</b> .....	22
<b>3.2. Metodologia</b> .....	22
<b>3.2.1. Base de dados</b> .....	22
<b>3.2.2. Indicadores de confiabilidade</b> .....	24
<b>3.2. Análise de confiabilidade e discussão dos resultados</b> .....	29
<b>3.2.1. Equivalência (concordância)</b> .....	29
<b>3.2.2. Estabilidade</b> .....	35
<b>3.2.3. Consistência interna</b> .....	39
<b>3.3. Considerações finais</b> .....	44
<b>4. ENSAIO II – UMA ANÁLISE DO EFEITO DAS CARACTERÍSTICAS DOS OBSERVADORES NAS MEDIDAS DE CONFIABILIDADE DA EAPI CONTÍNUA</b> .....	46
<b>4.1. Introdução</b> .....	46
<b>4.2. Metodologia</b> .....	47
<b>4.2.1. Base de dados</b> .....	47
<b>4.2.2. Procedimentos estatísticos</b> .....	50
<b>4.3. Resultados</b> .....	51
<b>4.4. Considerações finais</b> .....	55
<b>5. LIMITAÇÕES</b> .....	55
<b>6. CONCLUSÃO</b> .....	56
<b>REFERÊNCIAS</b> .....	58
<b>APÊNDICE A</b> .....	64
<b>APÊNDICE B</b> .....	72

## 1. INTRODUÇÃO GERAL

Investir nos primeiros anos de vida das crianças além de uma garantia de um direito fundamental, que é assegurado pelo Artigo 227 da Constituição Federal de 1988 (CF/88) e ratificado por meio do Artigo 4 do Estatuto da Criança e do Adolescente (ECA) que caracteriza as crianças e os jovens como um grupo que deve ter preferência na formulação e execução de políticas sociais públicas, também apresenta retornos futuros mais substanciais do que investimentos realizados em fases posteriores (Carneiro e Heckman, 2003). Esse período crucial na determinação dos resultados econômicos de um indivíduo, tais como o desempenho educacional e acadêmico (Heckman, 2006), habilidades socioemocionais (Jones; Bouffard, 2012), participação na força de trabalho (Carneiro e Heckman, 2003) e redução dos custos sociais (Reynolds *et al.*, 2001; Schweinhart *et al.*, 2005). Dessa forma, é importante compreender os aspectos que compõem essa fase de desenvolvimento.

As políticas públicas voltadas para aprimorar a oferta de educação infantil têm demonstrado um impacto significativo. Crianças que foram atendidas na creche possuem maiores chances de atingir níveis mais altos de alfabetização (Shirasu *et al.*, 2022), melhor desempenho em exames de leitura, matemática e ciências se comparadas tanto com as que foram atendidas somente na pré-escola como as que não receberam educação infantil (Irfi; Pereira, 2022) e maiores níveis de desenvolvimento das habilidades sociais e progresso escolar (Camilli *et al.*, 2010). Outros fatores afetados por políticas de melhoria da educação infantil são a redução das taxas de evasão escolar e de criminalidade (Reynolds *et al.*, 2002) e a redução de fatores de risco para doenças físicas e mentais (Campbell *et al.*, 2014).

Considerando os resultados observados e a expansão do acesso à educação infantil, juntamente com a importância crucial dessa fase para a formação e desenvolvimento das crianças, uma questão fundamental é se a qualidade da oferta é suficientemente elevada para garantir o desenvolvimento e a aprendizagem saudável das crianças (Marope; Kaga, 2015; Yoshikawa *et al.*, 2018). Com isso, se faz necessário analisar e avaliar a qualidade da educação infantil, a qual possui o protagonismo nas discussões a respeito de políticas públicas da primeira infância. Ademais, a aprendizagem de qualidade é um direito da criança, segundo a BNCC (Brasil, 2018).

Neste sentido, o uso de instrumentos que possam verificar a qualidade da educação ofertada tem sido discutido na literatura nacional e internacional (Shirasu *et al.*,

2021; Vitiello *et al.*, 2018), bem como tem sido expresso como metas do Plano Nacional de Educação<sup>1</sup> (PNE) para as políticas educacionais brasileiras, a fim de aumentar a taxa de matrícula de crianças na creche (até 50%) e na pré-escola (100%) até 2024.

Apesar de não ser uma tarefa simples, visto que se trata de um conceito atrelado à cultura e em constante evolução (Castilho; Ogando; Gil, 2021), a qualidade na educação infantil pode ser definida como a oferta de um ambiente seguro, estimulante e propício ao desenvolvimento integral da criança, que promove seu bem-estar físico, emocional, social e cognitivo. Isso inclui aspectos como a presença de profissionais qualificados, uma estrutura física adequada, materiais pedagógicos diversificados, interações positivas entre adultos e crianças, currículo sensível às necessidades individuais e parceria eficaz com as famílias (Raikes *et al.*, 2023).

Yoshikawa *et al.* (2013) destacam a importância das interações sensíveis e responsivas entre adultos e crianças para o desenvolvimento socioemocional na primeira infância. A qualidade dos cuidados infantis não parentais está significativamente relacionada a um melhor desempenho cognitivo e acadêmico aos quinze anos, com efeitos positivos crescentes observados em níveis mais altos de qualidade (Vandell *et al.*, 2010). Ainda de acordo com esses mesmos autores, os cuidados infantis de alta qualidade na primeira infância também estão associados positivamente com níveis adequados de saúde mental.

Além disso, Maia e Williams (2005) ressaltam a necessidade de ambientes físicos seguros e saudáveis na promoção do desenvolvimento infantil. Assim, para aumentar a eficácia da educação infantil, 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).

Campos *et al.* (2011) encontraram resultados que mostraram que as creches e pré-escolas brasileiras apresentaram, em média, níveis de qualidade insatisfatórios. 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 (Gullo, 2013; Yoshikawa *et al.*, 2018).

A coleta de dados sobre a qualidade dos ambientes de Primeira Infância desempenha um papel crucial na orientação dos sistemas de educação infantil (EI),

---

<sup>1</sup> Disponível em: <https://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>

oferecendo um feedback valioso aos formuladores de políticas e aos profissionais de EI sobre os níveis gerais de excelência e as áreas prioritárias para investimento visando o aprimoramento da qualidade (Thorpe *et al.*, 2022; Zaslow *et al.*, 2011).

A iniciativa *Measuring Early Learning Quality & Outcomes* (MELQO) foi criada para aprimorar a qualidade dos dados da EI e contribuir com o desenvolvimento infantil, especificamente para países de baixa e média renda (UNESCO, 2017). No Brasil, uma parte do MELQO foi adaptado por meio de um grupo de trabalho, e isso resultou na “Escala de Avaliação dos Ambientes de Aprendizagens dedicados à Primeira Infância” (EAPI)<sup>2</sup> (Fuzzari *et al.*, 2022). Pesquisas que integram a avaliação externa da qualidade de instituições de educação infantil com a observação do desenvolvimento das crianças podem proporcionar avanços no entendimento desses processos no país (Campos, 2020).

Medidas de qualidade na educação infantil contribuem para o fomento de políticas públicas e aperfeiçoamento de práticas, mas poucos estudos examinaram a adaptação e as propriedades psicométricas dessas medidas. Além disso, instrumentos de medição podem enfrentar problemas relacionados a sua validade e confiabilidade (Pilatti *et al.*, 2010).

Nesse sentido, faz-se necessário a discussão a respeito da qualidade dos instrumentos de coleta de dados. Nesse trabalho, será abordado o uso da EAPI, que será descrito com mais detalhes na próxima seção. A coleta de dados por meio da EAPI é realizada por aplicadores treinados, que se deslocam até as unidades educacionais e passam cerca de três horas e meia acompanhando uma turma de EI.

Durante a visita, a pontuação da maioria dos itens é com base no julgamento de práticas e ambientes, desse modo, alguns fatores podem afetar os observadores individualmente durante a aplicação do instrumento. Esses fatores podem influenciar a consistência e a precisão de suas respostas e os principais motivos são o viés de resposta (Paulhus, 2002), a variabilidade individual (Ones; Viswesvaran; Reiss, 2007), experiência e treinamento (Pianta; La Paro; Hamre, 2008), efeitos de fadiga (Dohrenwend *et al.*, 1980) e desatenção e carga cognitiva (Sweller, 1994).

Este trabalho está dividido em dois ensaios, o primeiro analisará a confiabilidade da aplicação da EAPI no formato contínuo de cinco aplicações por meio da estabilidade, consistência interna e equivalência (Souza *et al.*, 2017; Pilatti *et al.*, 2010; Downer *et al.*, 2010). E o segundo busca inferir sobre a disposição dos aplicadores e a quantidade de

---

<sup>2</sup> Disponível em: <https://observatorioei.org.br>

reaplicações utilizando características individuais e informações da coleta de campo como variáveis explicativas para a confiabilidade (Pianta; La Paro; Hamre, 2008; Downer *et al.*, 2010).

Vale ressaltar que a expansão da aplicação do instrumento para cinco dias na semana, ao invés de apenas um dia, teve como um dos objetivos, buscar um resultado mais robusto em termos das práticas efetivas, e tentar capturar o maior número possível de momentos de qualidade oportunizado às crianças, considerando que pode haver alguma programação de atividades que em dias específicos.

As duas primeiras seções serão de caráter introdutório. Além desta introdução, será apresentado a EAPI. Ademais, serão discutidos os ensaios nas seções seguintes, nelas estarão incluídas suas respectivas metodologias e discussões dos resultados encontrados. Por fim, algumas considerações finais para concluir o texto.

## **2. ESCALA DE AVALIAÇÃO DOS AMBIENTES DE APRENDIZAGENS DEDICADOS À PRIMEIRA INFÂNCIA (EAPI)**

A EAPI foi elaborada pelo Laboratório de Estudos e Pesquisas em Economia Social (LEPES/USP), com parceria da Fundação Maria Cecília Souto Vidigal (FMCSV) e tem por objetivo apoiar estratégias de avaliação que informem sobre a qualidade dos ambientes e das experiências ofertadas e vivenciadas pelas crianças na Educação Infantil, compreendendo as etapas de creche e pré-escola (Fuzzari *et al.*, 2022).

A escala tem como base o módulo MELE (*Measuring Early Learning Environment*, no original em inglês, ou Avaliando Ambientes de Educação Infantil, em tradução livre), integrante do instrumento MELQO que foi originalmente desenvolvido pelo Banco Mundial, UNICEF e UNESCO para fomentar a mensuração da qualidade da Educação Infantil (UNESCO, 2017).

### **2.1 Measuring Early Learning Environment (MELE)**

O MELQO é composto por dois módulos, o MODEL (*Measure of Development and Early Learning*) e o MELE. Um desses módulos é voltado para medir o desenvolvimento e aprendizagem das crianças (MODEL), enquanto o outro se concentra em avaliar a qualidade dos ambientes de aprendizagem (MELE) na pré-escola e séries iniciais, reunindo sete grandes áreas: brincadeira; pedagogia; interações; ambiente;

formação da equipe escolar; comprometimento de pais e comunidades; e inclusão (UNESCO, 2017).

Os dois módulos foram testados em alguns países (Colômbia, Quênia, Nicarágua e Tanzânia) e a partir de resultados preliminares foi possível confirmar a confiabilidade do MELQO com testes “interobservadores” (diferentes aplicadores são capazes de obter pontuações semelhantes nos itens) e de consistência interna (os itens que medem o mesmo constructo se correlacionam entre si). A validação dos instrumentos, por meio das propriedades de constructo (os itens medem o que dizem medir), relevância intercultural (os itens funcionam de forma semelhante em diferentes culturas) e simultaneidade (os itens se correlacionam com outros conjuntos de itens estabelecidos), entretanto, até o momento a predição (o instrumento prevê o desenvolvimento futuro das crianças) ainda não havia sido estabelecida (World Bank, 2017).

Esses resultados foram confirmados em aplicações subsequentes do MELQO na Colômbia e na Tanzânia, o MODEL obteve propriedades psicométricas aceitáveis (Raikes *et al.*, 2019), já o MELE necessitou realizar adaptações significativas para adequar-se a cada contexto, resultando em associações modestas com o desenvolvimento infantil (Maldonado *et al.*, 2022).

Além das propriedades psicométricas das medidas, várias lições surgiram das experiências dos países que utilizaram o MELQO. Essas lições incluem reflexões sobre os recursos necessários para adaptar e implementar as ferramentas em grandes grupos de crianças, os desafios em treinar os coletores de dados para confiabilidade quando os conceitos descritos nas medidas não são familiares para eles e a importância de alinhar as ferramentas com as prioridades e perspectivas locais sobre o desenvolvimento infantil e a qualidade dos ambientes de aprendizagem (Raikes, 2023).

Os dados do MELQO foram utilizados para abordar duas questões fundamentais na formulação de políticas. Em primeiro lugar, eles destacaram as disparidades no desenvolvimento infantil, influências na aprendizagem nos primeiros anos e na qualidade dos ambientes de aprendizagem das crianças. Em segundo lugar, apesar da falta de comparabilidade internacional dos resultados das medidas, os dados descritivos dos estudos do MELQO podem identificar áreas específicas e acionáveis para melhorias nos sistemas de Educação Infantil.

Su *et al.*, (2021) utilizaram o MELE para investigar a associação entre a qualidade da pré-escola (crianças de 4 e 5 anos) e o desenvolvimento das crianças (medido pelo *East Asia-Pacific Early Child Development Scales*) na província de Guizhou e em

Xangai. Além disso, eles exploraram os impactos da localização geográfica nessa associação e encontraram que a qualidade das unidades educacionais foi superior em Xangai, menor em Guizhou e ainda menor em áreas rurais desta última região.

A qualidade pré-escolar apresentou uma associação positiva com os resultados das crianças, sendo que análises indicaram uma correlação mais fraca em áreas urbanas de Guizhou e Xangai. A partir disso, os autores sugerem que ambientes de qualidade na pré-escola podem oferecer benefícios significativos para crianças em ambientes rurais e desfavorecidos.

Em outro contexto, um estudo que abrange a aplicação do MELQO no município de Sobral (CE), permitiu Lima (2022) encontrar evidências que as expectativas parentais são influenciadas pelo contexto socioeconômico familiar e escolar.

## **2.2 Elaboração da EAPI e a EAPI contínua**

Para alcançar a versão atual da EAPI foi realizado um processo de alinhamento à Base Nacional Comum Curricular (Brasil, 2018). Ela é estruturada para possibilitar a observação de oportunidades de aprendizagens nas diferentes dimensões que compõem a EI e busca investigar o que é ofertado em termos de condições dos ambientes de aprendizagens (infraestrutura, currículo, práticas pedagógicas, interações, diversidades, segurança, alimentação, equipe e gestão).

A EAPI é um instrumento que tem como objetivo entender e expressar o nível da qualidade do ensino na primeira infância por meio de uma medida numérica mais precisa possível, obtida por meio de observações sistemáticas de parcelas do atendimento a crianças nas unidades de educação. Parte dos itens podem ser verificados dentro da rotina das unidades educacionais sendo, alguns pontuados a partir de uma escala de 1 a 4 pontos, em que pontuações mais altas significam maior qualidade, e mais baixas revelam menor qualidade. Outros itens podem ser pontuados a partir de informações coletadas com a(o) professora(or) responsável pela turma.

Os aplicadores passam por um treinamento especializado para compreender e executar os instrumentos da EAPI<sup>3</sup> e uma prova de certificação. O segundo ensaio irá explorar as informações acerca da formação dos aplicadores, bem como suas características individuais, desempenho e conhecimento sobre o desenvolvimento infantil. Com isso, irá inferir sobre a atuação deles em campo utilizando os indicadores

---

<sup>3</sup> Composto por duas entrevistas (uma voltada para professores e uma para diretores) e um roteiro de observação direta.

de confiança e dados da coordenação de campo, como por exemplo a quantidade de vezes que aplicaram o instrumento, total e por semana, tempo entre aplicações e distância até a unidade educacional.

Em 2021, a EAPI foi aplicada em 12 municípios do Brasil, dentro de um Estudo Nacional sobre a Qualidade da Educação Infantil<sup>4</sup>. O Mapa 1 mostra as doze cidades escolhidas para realização da coleta. O objetivo do estudo era colher dados de todas as cinco regiões brasileiras, a fim de mostrar um cenário de escala nacional. A coleta foi amostral nessas cidades e a representatividade foi estipulada por meio de quantidades de turmas correspondente as faixas populacionais.

Nessa fase, a amostra foi composta por um município com 600 turmas, dois municípios com 400 turmas cada, três com 320 turmas cada, três municípios com 240 turmas cada, dois municípios com 180 turmas cada e um com 120 turmas, totalizando 3.560 turmas onde foram aplicados os instrumentos da EAPI.

**Figura 1** - Municípios participantes do estudo nacional (2021).



Fonte: Estudo nacional de qualidade da educação infantil (2021).

Os resultados contidos no relatório são obtidos utilizando uma metodologia que consiste em organizar os itens da EAPI em seis dimensões, são elas: alimentação; currículo, interações e práticas pedagógicas; diversidade funcional; equipe e gestão;

<sup>4</sup> Ver <https://www.fmcsv.org.br/pt-BR/biblioteca/estudo-nacional-qualidade-educacao-infantil-2022/>

infraestrutura; e segurança. Como pode ser observado no Quadro 1, essas dimensões são compostas por subgrupos formados por itens específicos que avaliam parâmetros de qualidade na educação infantil.

Por exemplo, “Estrutura predial” contém todos os itens referentes ao espaço físico (iluminação, ventilação, janelas, bancadas, espelho, parques, bebedouro, etc.) da unidade educacional, e juntamente com os itens de “Materiais” e “Equipamentos” formam a dimensão “Infraestrutura”. Cada dimensão possui itens do roteiro de observação, enquanto as entrevistas do instrumento contribuem para a dimensão “Equipe e gestão” e algumas seções das demais.

**Quadro 1 – Dimensões da EAPI.**

<b>Dimensão</b>	<b>Seções</b>	<b>Número de itens</b>
Alimentação	- Gestão	5
	- Infraestrutura	4
	- Práticas	4
Currículo, interações e práticas pedagógicas	- Organização dos tempos, espaços e materiais	19
	- Planejamento e currículo	8
	- Práticas de ampliação do repertório por meio do brincar	12
	- Práticas para acolhimento e gestão de conflitos	7
Diversidade funcional	- Práticas para o cuidado de si, o bem-estar e a saúde	9
	-	22
Equipe e gestão	- Equipe	21
	- Gestão	20
Infraestrutura	- Equipamentos	19
	- Estrutura predial	23
	- Materiais	11
Segurança	-	15

Fonte: Estudo nacional de qualidade da educação infantil (2022).

Em outra fase<sup>5</sup>, dando continuidade ao estudo da qualidade da educação infantil, foi elaborado um novo formato de aplicação do instrumento. Denominado de “EAPI contínua”, esse modelo consiste na aplicação do roteiro de observação em uma mesma turma nos cinco dias da semana, não necessariamente de forma sequencial. As entrevistas de professores e diretores foram aplicadas apenas uma vez para cada turma e unidade educacional, respectivamente. O propósito dessa iniciativa envolve algumas discussões das limitações inerentes ao instrumento, tais como a restrição da aplicação convencional

<sup>5</sup> Antes desse período, houve outras aplicações em 12 municípios do Ceará, em Patos (PB), Barretos (SP), entre outros.

em capturar apenas uma visão parcial da experiência diária, sujeita a influência de fatores externos ao atendimento das crianças e dias atípicos.

O uso contínuo do instrumento surge como forma de tentar captar momentos de oportunidades ao longo da semana, dado que algumas atividades ocorrem em dias específicos. Também busca ter uma visão mais detalhada da documentação pedagógica diária por meio dos itens de planejamento e registros de aprendizagens e produções das crianças, com base no princípio 4.4 (Observação, planejamento, documentação e reflexão das práticas pedagógicas e das aprendizagens das crianças) dos Parâmetros Nacionais de Qualidade da Educação Infantil<sup>6</sup>.

Além disso, testar a consistência da EAPI de uma aplicação em relação a cinco (ou menos), buscar uma quantidade ótima de observações, discutir sobre a prática diferenciada por parte de professoras de referência, realizar um acompanhamento da rotina, verificar se a observação única retrata bem a rede ou se por ser um evento único não retrata de forma fidedigna a realidade e trazer o olhar de mais de um observador sobre a situação.

A disposição de visitação foi dividida em três formatos diferentes:

- (i) Apenas um observador acompanha a turma durante os cinco dias da semana;
- (ii) No segundo, dois observadores se alternam entre os dias, com um deles indo em uma segunda, quarta e sexta, e o outro numa terça e quinta; e
- (iii) No terceiro, cinco observadores diferentes foram nas turmas aplicar o instrumento, um em cada dia.

Buscou-se manter o equilíbrio entre os três tipos de observação propostos, com aproximadamente um terço das turmas em cada formato. Ao todo, a coleta durou 17 semanas, para isso, 30 aplicadores foram a campo, com uma média de 13 observadores em cada semana.

Segundo dados de maio de 2022, fornecidos pela rede municipal de ensino de onde foi aplicada a EAPI contínua, havia um total de 412 turmas na educação infantil. A partir desses dados, foi realizada uma amostragem para definir as turmas que seriam observadas. Considerando uma amostra que pudesse ser representativa em nível de rede municipal, foi realizada uma amostragem aleatória estratificada com 147 turmas, entre unidades de creche (apenas de equipamentos do pró-infância, os Centros de Educação

---

<sup>6</sup> Ver: <http://portal.mec.gov.br/docman/2020/141451-public-mec-web-isbn-2019-003/file>

Infantil) e pré-escola (apenas de unidades com etapas integradas, denominadas como patrimoniais).

O instrumento de observação direta foi aplicado cinco vezes em cada turma, enquanto as entrevistas de professores(as) e diretores(as) foram aplicadas uma vez para cada turma e unidade escolar, respectivamente.

### **3. ENSAIO I: UMA ANÁLISE DA CONFIABILIDADE DA APLICAÇÃO CONTÍNUA DA ESCALA DE AVALIAÇÃO DOS AMBIENTE DE APRENDIZAGEM DEDICADOS À PRIMEIRA INFÂNCIA (EAPI)**

#### **3.1. Introdução**

Quando se trata de medir atributos em pesquisas, práticas ou avaliações, a confiabilidade desempenha um papel fundamental. Ela se refere ao grau de consistência com que os itens do instrumento medem o atributo proposto, livres de erro de medida. Além disso, a confiabilidade também está relacionada ao grau em que o instrumento permite a reprodução e a obtenção de resultados consistentes, quando aplicado em diferentes ocasiões (Echevarría-Guanilo; Gonçalves; Romanoski, 2018).

Para garantir resultados confiáveis e válidos, é essencial compreender os conceitos de confiabilidade e conhecer os diferentes métodos de avaliação. A confiabilidade é relevante em estudos de validação de instrumentos, onde a consistência e a estabilidade das medidas são essenciais para a qualidade metodológica (Alexandre *et al.*, 2013; Salmond, 2008).

Os instrumentos precisam gerar dados confiáveis, válidos e que possam ser interpretados. Da mesma forma, as medidas devem apresentar resultados com rigor científico. Apesar de não ser uma regra universal, é consensual na literatura que a qualidade dos resultados dessas medidas depende, em grande parte, da confiabilidade e da validade dos instrumentos (Cook; Beckman, 2006).

De acordo com Raikes *et al.* (2019), as análises do MELE indicaram uma consistência interna adequada, exceto para o desenvolvimento das crianças e características dos professores. Esses resultados serão o ponto de partida para realizar uma análise semelhante para EAPI.

A próxima subseção apresenta a metodologia empregada nesse ensaio, iniciando com a caracterização da base de dados e aprofundando nas definições dos indicadores de confiabilidade. Em seguida, serão apresentados os resultados.

#### **3.2. Metodologia**

##### **3.2.1. Base de dados**

As observações e coletas de dados da EAPI Contínua foram realizadas em 16 Centros de Educação Infantil (CEI) com turmas com crianças de 2 a 3 anos e 11 meses e

35 Escolas Municipais de Tempo Parcial (EMTP) com turmas de crianças de 4 a 5 anos e 11 meses. Ao todo foram realizadas 735 observações, sendo 5 dias em cada uma das 147 turmas distribuídas entre as unidades sorteadas. No entanto, após um tratamento de valores ausentes e de concatenação entre a observação direta e as entrevistas, a base de dados ficou composta com informações de 144 turmas.

Na Tabela 1 estão algumas estatísticas descritivas a respeito dos dados coletados, como a quantidade de observadores que foram a campo, o número total, médio, máximo e mínimo de crianças observadas, agregado e discriminado por dia da semana e sexo, além da quantidade professoras. Não é possível saber a quantidade total exata de crianças e professoras observadas, pois alguns indivíduos se repetem em alguns dias e outros não. Em resumo, foram observadas em média 3 turmas por unidade escolar, com 19,5 crianças matriculadas e 13 presentes no dia da observação, em média, em cada uma. O tamanho das turmas selecionadas variou de 14 a 22 matrículas com um total de 2.869 crianças somadas.

**Tabela 1** – Estatísticas descritivas da base de dados.

	<b>N</b>	<b>Média</b>		<b>Min</b>		<b>Max</b>		
Observadores	30	-	-	-	-	-	-	
Turmas**	144	3,0		1		6		
Unidades Escolar	51	-		-		-		
Crianças matriculadas*	2869	19,5		14		22		
Crianças presentes*	-	13,0		-		-		
1º dia	1822	12,6		5		20		
2º dia	1924	13,4		5		20		
3º dia	1932	13,4		6		21		
4º dia	1863	12,9		3		21		
5º dia	1830	12,7		5		20		
Crianças presentes*		<b>(1)</b>	<b>(2)</b>	<b>(1)</b>	<b>(2)</b>	<b>(1)</b>	<b>(2)</b>	
Meninas (1) / Meninos (2)	-	-	6,3	6,7	-	-	-	-
1º dia	881	941	6,1	6,5	0	1	12	15
2º dia	925	1006	6,4	7,0	1	1	12	14
3º dia	948	984	6,6	6,8	1	0	13	14
4º dia	897	966	6,2	6,7	1	1	13	15
5º dia	887	943	6,2	6,5	2	0	12	15
Professoras	-	-	-	-	-	-	-	
1º dia	169	-	-	-	-	-	-	
2º dia	163	-	-	-	-	-	-	
3º dia	158	-	-	-	-	-	-	
4º dia	162	-	-	-	-	-	-	
5º dia	159	-	-	-	-	-	-	

\*Por turma

\*\*Por unidade educacional

Fonte: Elaborado a partir dos dados da EAPI Contínua.

A quantidade de meninos foi maior que a de meninas em todos os dias de observação, tanto para o total, como para a média e o valor máximo. Contudo, houve turmas em dois dias que não tiveram presença de meninos e apenas um dia com turmas que não tiveram meninas presentes. O número de professoras na sala de referência foi maior no primeiro dia de observação em relação aos demais, e reduziu em 6,5% até o terceiro dia. A média foi de 162,2 professoras presentes por dia durante as cinco visitas dos aplicadores às turmas.

### ***3.2.2. Indicadores de confiabilidade***

Um instrumento devidamente elaborado deve levar em consideração dois aspectos muito importantes: sua validade e sua confiabilidade. A Figura 2 esquematiza as relações potenciais entre confiabilidade e validade. O primeiro alvo não exibe nem confiabilidade nem validade, com pontos dispersos apenas na parte lateral. No segundo alvo, os lances são consistentes entre si, atingindo pontos próximos uns aos outros, embora não atinjam o centro do alvo, resultando em falta de validade.

O terceiro alvo pode ser considerado válido, embora não confiável uma vez que os pontos atingidos não se concentraram em um ponto específico, mas se espalharam por uma região fora do centro. O quarto exemplo exemplifica a combinação ideal de confiabilidade e validade, onde os lances atingem consistentemente o alvo pretendido, centrando-se no ponto desejado. Essas relações também podem ser aplicadas à avaliação das propriedades de medida dos instrumentos.

**Figura 2** – Relações entre confiabilidade e validade.



Fonte: Elaborado pelo autor

Neste ensaio será investigada a confiabilidade da EAPI, dado o seu formato de aplicação contínua. Ela se refere à capacidade de um instrumento em produzir resultados consistentes ao longo do tempo, entre diferentes observadores e em diferentes contextos. Existem três critérios principais de confiabilidade: estabilidade, consistência interna e equivalência (Souza *et al.*, 2017). Cada critério possui métodos estatísticos específicos para sua avaliação.

### **Estabilidade**

A estabilidade refere-se à consistência dos resultados obtidos em diferentes momentos no tempo. Pode ser avaliada usando o método de teste-reteste, no qual a mesma medida é aplicada em dois ou mais períodos distintos, a EAPI contínua permite essa avaliação devido a visita de um observador nos cinco dias da semana na turma. A confiabilidade do teste-reteste pode ser estimada usando o coeficiente de correlação intraclassa (ICC), que leva em consideração os erros de medida (Koo; Li, 2016).

O ICC é utilizado em análises de dados de medidas repetidas, em que as mesmas unidades são medidas em diferentes ocasiões. Pode ser calculado para diferentes tipos de dados, incluindo dados contínuos, categóricos e ordinais. Existem diferentes fórmulas para o cálculo do ICC, a depender do tipo dos dados e do design do estudo. Aqui será

utilizada a fórmula de Shrout e Fleiss (1979), o  $ICC(3, k)$ , também chamado de modelo de duas vias mistos. O  $ICC(3, k)$  é uma média ponderada dos coeficientes  $ICC(1, k)$ , modelo de uma via aleatório, e o  $ICC(2, k)$ , modelo de duas vias aleatório, e é obtido da seguinte forma:

$$ICC(3, k) = \frac{MSA - MSE}{MSA + (k - 1)MSE + \frac{(MSA - MSE)}{n}} \quad (1)$$

Onde:  $MSA$  é a média das variâncias das medidas de cada avaliador;  $MSE$  é a média das variâncias dentro dos avaliadores;  $k$  é o número de avaliadores; e  $n$  é o número total de medidas.

O ICC varia de 0 a 1, sendo que valores mais altos indicam maior concordância entre as medidas. Um valor igual a 0 indica ausência de concordância, enquanto um valor igual a 1 indica concordância perfeita. Quando a variabilidade entre os avaliadores é muito maior do que a variância dentro dos avaliadores, o  $ICC(3, k)$  se aproxima do  $ICC(1, k)$ , indicando que os efeitos específicos de cada avaliador são menos importantes. Por outro lado, quando a variância entre os avaliadores é muito menor do que a variância dentro dos avaliadores, o  $ICC(3, k)$  se aproxima do  $ICC(2, k)$ , indicando que os efeitos específicos de cada avaliador são mais importantes.

### **Consistência interna**

A consistência interna indica se todas as partes de um instrumento medem a mesma característica. O coeficiente alfa de Cronbach é comumente utilizado para essa análise, refletindo a covariância entre os itens de uma escala, no entanto, é importante ressaltar que os valores deste coeficiente são fortemente influenciados pelo número de itens. É uma medida estatística que varia de 0 a 1, sendo que se estiver próximo de 1 indica que a escala tem alta consistência interna, ou seja, que os itens estão altamente relacionados e medem o mesmo construto. E baixa consistência, caso contrário.

Para estimar o coeficiente alfa de Cronbach, é necessário ter um conjunto de itens que avaliem um mesmo construto ou dimensão, esse indicador, portanto, será calculado para as seis dimensões da EAPI. Avaliando as pontuações atribuídas por cada aplicador, é factível calcular a correlação entre elas. Quanto mais elevada for essa correlação, maior será o valor associado ao coeficiente. O cálculo do alfa de Cronbach leva em consideração

a variância total da escala, a variância dos escores de cada item e a média das correlações entre os itens e se dá pela seguinte forma:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2}\right) \quad (2)$$

Onde:  $k$  é o número de itens;  $s_i^2$  é a variância de cada item; e  $s_t^2$  é a variância de todos os itens.

O coeficiente  $\alpha$  pode ser determinado em duas fases: inicialmente, é preciso calcular a variância de cada coluna  $i$  da matriz (3), e em seguida, somar todas essas variâncias. Na segunda fase, deve-se obter a soma total das avaliações de cada avaliador (elementos da última coluna da matriz de julgamentos dos  $k$  itens por  $n$  avaliadores) e, em seguida, calcular a variância dessas somas.

$$\text{matriz de julgamentos} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \left[ \begin{array}{c} \sum x_{1k} \\ \sum x_{2k} \\ \vdots \\ \sum x_{nk} \end{array} \right] \\ x_{21} & x_{22} & \dots & x_{2k} & \\ \vdots & \vdots & \ddots & \vdots & \\ x_{n1} & x_{n2} & \dots & x_{nk} & \end{bmatrix} \quad (3)$$

Como o instrumento analisado possui diferentes escalas de medição das variáveis, neste estudo será utilizado o coeficiente  $\alpha$  de Cronbach padronizado. Ele é obtido por meio de uma transformação das pontuações originais das variáveis. A fórmula para calcular o alfa padronizado é a seguinte:

$$\alpha_p = \frac{k\bar{r}}{1 + (k-1)\bar{r}} \quad (4)$$

Onde:  $k$  é o número de itens e  $\bar{r}$  é a média das correlações item-total padronizadas.

O valor de  $\bar{r}$  é determinado pela média de  $r_i$  calculada para todos os itens no teste. A correlação item-total padronizada ( $r_i$ ) é calculada para cada item individual em relação à soma total dos escores de todos os outros itens da seguinte forma:

$$r_i = \frac{\text{Cov}(\text{Item}_i, \text{Total dos outros itens})}{\sqrt{\text{Var}[\text{Item}_i + \text{Var}(\text{Total dos outros itens})]}} \quad (5)$$

### Equivalência

E a equivalência diz respeito a concordância entre observadores em relação aos escores de um instrumento. A confiabilidade “interobservadores” é a forma mais comum

de avaliar a equivalência e envolve a participação de dois ou mais avaliadores independentes preenchendo o mesmo instrumento. O coeficiente Kappa é uma medida utilizada para medir a concordância entre observadores que classificam o mesmo objeto, fenômeno ou evento. É calculado pela seguinte fórmula:

$$\kappa = \frac{(P_{obs} - P_{exp})}{(1 - P_{exp})} \quad (6)$$

Onde:  $P_{obs}$  é a proporção observada de concordância entre os dois avaliadores (quantas vezes os avaliadores concordaram); e  $P_{exp}$  é a proporção esperada de concordância entre os dois avaliadores, assumindo que a concordância é devida ao acaso. O resultado do coeficiente Kappa será um número entre -1 e 1, onde valores maiores indicam maior concordância entre os observadores.

No Quadro 2 é mostrado um resumo das interpretações mais comumente utilizadas na literatura para os indicadores usados nesta avaliação. Como o instrumento da EAPI é dividido em 6 dimensões, serão calculados coeficientes de confiabilidade para cada uma separadamente. Assim, utiliza-se de uma classificação validada dos itens e espera-se chegar a resultados que possuem explicação mais simples e detalhada.

**Quadro 2 – Interpretação par os coeficientes de confiabilidade.**

	ICC	Alfa	Kappa
Insignificante	-	Menor que 0,30	Menor que 0,00
Fraco	Menor que 0,50	De 0,31 a 0,60	De 0,01 a 0,20
Razoável	-	-	De 0,21 a 0,40
Moderado	De 0,51 a 0,75	De 0,61 a 0,70	De 0,41 a 0,60
Bom	De 0,76 a 0,90	De 0,71 a 0,80	De 0,61 a 0,80
Excelente	De 0,91 a 1,00	De 0,81 a 0,90	De 0,81 a 1,00
Perfeito ou redundante	-	Maior que 0,91	-

Fonte: Elaborado pelo autor a partir de Freitas e Rodrigues (2005), Cicchetti (1994) e Landis e Koch (1977).

A estimação dos coeficientes a nível da turma irá contar apenas com itens do instrumento de observação direta, pois as entrevistas foram realizadas apenas uma vez<sup>7</sup>, desse modo, a dimensão de “Equipe e gestão” não será avaliada por ser composta quase inteiramente por esses itens. No entanto, será realizado um exercício que calcula os coeficientes considerando dados agregados, nisso cada turma contará uma aplicação de

<sup>7</sup> Nesse cenário, após concatenar os 3 instrumentos, os dados das entrevistas apareceriam repetidos cinco vezes (para cada observação direta), não possuindo variabilidade.

cada instrumento, como é convencionalmente utilizada a EAPI. Assim, todas as dimensões poderão ser analisadas.

### 3.2. Análise de confiabilidade e discussão dos resultados

Nesta seção serão discutidos os resultados obtidos a partir da metodologia.

#### 3.2.1. Equivalência (concordância)

O coeficiente Kappa<sup>8</sup> foi calculado para todas as dimensões da EAPI a nível da rede municipal, segmentado pela faixa etária da educação infantil. Para este resultado, foram avaliadas as 144 aplicações do instrumento nas turmas de cada dia separadamente. É possível observar pela Tabela 2 que as dimensões assumem valores próximos nas cinco coletas diferentes. Quando a amostra é dividida considerando as faixas de creche e pré-escola, o Kappa se torna maior, devido aos grupos possuírem características semelhantes.

**Tabela 2** – Coeficiente Kappa a nível de rede por dia observado e por faixa etária.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Equipe e gestão	Infraestrutura	Segurança
<i>Toda a amostra</i>						
1º dia	0,50	0,31	0,35	0,35	0,39	0,37
2º dia	0,48	0,29	0,33	0,36	0,39	0,35
3º dia	0,47	0,28	0,34	0,35	0,39	0,34
4º dia	0,52	0,30	0,33	0,36	0,39	0,34
5º dia	0,47	0,29	0,33	0,36	0,39	0,33
<i>Creche</i>						
1º dia	0,51	0,36	0,51	0,36	0,44	0,54
2º dia	0,50	0,34	0,51	0,36	0,46	0,46
3º dia	0,51	0,33	0,57	0,36	0,46	0,47
4º dia	0,55	0,32	0,56	0,37	0,45	0,46
5º dia	0,52	0,31	0,52	0,36	0,45	0,44
<i>Pré-escola</i>						
1º dia	0,49	0,31	0,40	0,36	0,41	0,35
2º dia	0,49	0,29	0,38	0,36	0,40	0,33
3º dia	0,46	0,29	0,39	0,36	0,41	0,32
4º dia	0,52	0,31	0,39	0,37	0,40	0,32
5º dia	0,47	0,30	0,37	0,36	0,41	0,31

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

De acordo com a classificação mostrada no Quadro 2 quando se observa toda a rede de ensino, a dimensão “Alimentação” assume valores moderados de Kappa, enquanto as demais ficaram com coeficientes considerados apenas razoáveis, o que significa indícios de menor concordância entre os avaliadores.

Contudo, calcular o coeficiente Kappa para um grande número de aplicadores em relação ao número de itens, como é o caso em discussão, ocasiona desafios substanciais

<sup>8</sup> Todos os coeficientes Kappa dessa seção foram estimados com Intervalo de Confiança de 95%.

na análise da confiabilidade. Por exemplo, viés de concordância aleatória (Fleiss, Levin e Paik, 2013), assimetria nas margens da tabela de contingência, número insuficiente de categorias, complexidade na interpretação (devido à presença de muitas combinações possíveis de concordância e discordância) e amostras desequilibradas podem influenciar o Kappa (Feinstein e Cicchetti, 1990; Sim e Wright, 2005).

Os primeiros resultados da Tabela 2 tratam os dados como cinco aplicações convencionais distintas da EAPI, isso mostra que para esse formato, deve-se optar por uma abordagem diferente para determinar o Kappa. Uma possível alternativa é dividir a amostra por faixa etária da educação infantil, como explorado nas duas últimas seções dessa mesma tabela.

Após separar a amostra, para as observações realizadas na creche, as dimensões de “Alimentação”, “Diversidade funcional”, “Infraestrutura” e de “Segurança” atingem a categoria moderado. “Currículo, interações e práticas” e “Equipe e gestão” (essa dimensão é composta por 96% de itens referentes as entrevistas de professores e diretores, que foram realizadas apenas uma por turma, ou seja, os dados se repetem para cada turma e não necessariamente foram colhidos pelo mesmo aplicador da observação direta) permanecem com indicador razoável de concordância.

Para a pré-escola, apenas “Alimentação”, e “Infraestrutura” assumem valores considerados moderados. As demais permanecem razoáveis e não ocorre o mesmo visto para a creche, nesse caso, as dimensões “Diversidade funcional” e “Segurança” possuem menor concordância para a segunda faixa etária. Pelas características da amostra, sabe-se que esses grupos dividem também os tipos de unidades educacionais, em CEIs e EMTP, e considerando que essas duas dimensões possuem muitos itens referentes aos espaços e condições da UE, pode haver impacto das características individuais dessas categorias de instituições de educação na estimação.

Desse modo, além da faixa etária, agrupar tipos de instituições pode ser outra sugestão para determinar a equivalência dos observadores. Além disso, cabem outras análises que levem em consideração os aplicadores, como a experiência prévia, a formação específica e a familiaridade com o contexto da educação infantil.

Na principal análise de equivalência neste estudo, devido à aplicação contínua do instrumento, foi possível calcular o coeficiente Kappa a nível de turma observada. Dessa maneira, tornou-se viável avaliar o grau de concordância entre os observadores, considerando que o instrumento está sendo empregado para avaliar a qualidade dos ambientes em cada turma.

A Tabela 3 mostra como foi realizado o cálculo do Kappa para a dimensão “Infraestrutura” de uma das turmas da amostra. As colunas determinam a ordem de visita, na matriz triangular estão os coeficientes ( $m_{ij}$ ) para o par de observadores (i, j) de acordo com a equação (6). Na última linha está o Kappa medido em relação ao número de dias, ele é obtido pela média dos valores de  $m_{ij}$  acumulados até a coluna em questão, e o resultado em destaque é o Kappa da turma.

**Tabela 3** – Exemplo de matriz construída para cálculo do Coeficiente Kappa de uma turma.

	1º dia	2º dia	3º dia	4º dia	5º dia
Observador	A	B	C	D	E
A	-	0,70	0,51	0,44	0,59
B		-	0,70	0,44	0,71
C			-	0,51	0,59
D				-	0,48
E					-
(**)	-	0,70	0,64	0,55	<b>0,57(*)</b>

(\*) Coeficiente Kappa da turma utilizando toda a amostra;

(\*\*) Kappa por dia adicionado.

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

No 1º dia como foi apenas um observador, não há valor de comparação, a partir do 2º dia temos o Kappa entre A - B. No 3º dia é acrescentado os dados de observação do aplicador C e é obtido o Kappa entre A - C e A - B. Esses valores vão sendo agregados e acumulados a cada dia adicionado para obter um valor final para cada turma. A partir da última linha da Tabela 3 computado para todas as turmas, chega-se à Tabela 4, que mostra os resultados da confiabilidade “interobservadores” discriminado pelas dimensões. Com exceção de “Currículo, interações e práticas”, que apresentou nível de concordância moderado, as demais dimensões performaram um grau bom de concordância. A adição de uma observação a mais por turma não altera o nível de confiabilidade dos observadores e apresenta uma variação apenas na segunda casa decimal.

**Tabela 4** – Coeficiente Kappa a nível de turma por dimensão e dia adicional de observação.

	2º dia	3º dia	4º dia	5º dia
Alimentação	0,67	0,65	0,66	0,66
Currículo, interações e práticas	0,49	0,48	0,48	0,48
Diversidade funcional	0,64	0,65	0,65	0,65
Infraestrutura	0,65	0,64	0,63	0,63
Segurança	0,68	0,67	0,67	0,66

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

O coeficiente mais baixo para “Currículo, interações e práticas” pode ser explicado pela natureza específica dos itens do instrumento da EAPI que essa dimensão reúne. A baixa concordância pode ser resultado da complexidade ou da subjetividade dos critérios utilizados para avaliar esses aspectos específicos do instrumento. Itens que abordam nuances detalhadas, variedades de práticas ou interpretações diferentes podem levar a uma menor concordância entre os observadores, contribuindo para o coeficiente mais baixo observado nessa dimensão.

Considerando os subgrupos que compõe as dimensões, conforme visto no Quadro 1, pode-se aprofundar na análise dos elementos relacionados ao “Currículo, interações e práticas” e identificar possíveis discrepâncias específicas que contribuem para o resultado encontrado. Na Tabela 5, nota-se que os itens relacionados às “Práticas de ampliação do repertório por meio do brincar” e ao “Planejamento e currículo” geram menos concordância entre os observadores quando comparadas com as demais seções. Ambas apresentam nível razoável, enquanto o restante tem concordância moderada entre os observadores.

**Tabela 5** – Coeficiente Kappa a nível de turma por seções que compõe a dimensão currículo, interações e práticas.

	2º dia	3º dia	4º dia	5º dia
Organização dos tempos, espaços e materiais	0,47	0,47	0,47	0,47
Planejamento e currículo	0,35	0,37	0,36	0,36
Práticas de ampliação do repertório por meio do brincar	0,30	0,30	0,30	0,30
Práticas para acolhimento e gestão de conflitos	0,48	0,46	0,46	0,46
Práticas para o cuidado de si, o bem-estar e a saúde	0,47	0,45	0,46	0,46

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

Em relação aos itens que compõem essas seções (ver Tabela A1), há questões sobre planejamento das atividades, aprendizagens e produções das crianças com suas respectivas periodicidades, e utilização de documento curricular oficial na seção de planejamento e currículo. Especificamente nessa parte do questionário, o observador é instruído a preencher o instrumento mediante resposta da professora as perguntas, sendo que para marcar a presença desses itens, o aplicador deve tê-lo em posse de suas mãos. Os itens de periodicidade são condicionados a essa ação.

Já em relação às práticas de ampliação do repertório por meio do brincar, há itens que buscam medir a capacidade de as professoras utilizarem estratégias, com base em seus conhecimentos adquiridos, para oportunizar aprendizagens às crianças segundo as definições de aprendizagem baseada no brincar. Ambas essas seções encontram desafios

que podem fazer com que as interpretações do Kappa não sejam claras. Por exemplo, a decisão do observador pode estar condicionada a fatores que não são aleatórios, como a rotina e a professora que está com a turma.

A seção “Práticas de ampliação do repertório por meio do brincar” foi tratada, transformando as respostas aos itens relacionados as oportunidades de aprendizagem de duas formas diferentes. Na primeira<sup>9</sup>, foi levado em consideração apenas se o aplicador observou ou não a oportunidade referente ao item, na segunda<sup>10</sup>, além da abordagem anterior, foi diferenciado se a oportunidade de aprendizagem era de qualidade ou não.

Os observadores têm maior nível de concordância quando as opções de resposta são “houve” ou “não houve” a oportunidade de aprendizagem, do que quando há opções que qualificam a oportunidade (ver Tabela 6). Isso pode indicar que há desequilíbrios nas categorias ou que há fatores que aumentam a concordância esperada ao acaso e, conseqüentemente, reduz o valor do Kappa.

**Tabela 6** – Coeficiente Kappa a nível de turma: análise das seções “Planejamento e currículo” e “Práticas de ampliação do repertório por meio do brincar”,

	2º dia	3º dia	4º dia	5º dia
<b>Seção: Práticas de ampliação do repertório por meio do brincar (PARMA)</b>				
Presença de oportunidade de aprendizagem	0,45	0,45	0,45	0,44
Presença e qualidade da oportunidade de aprendizagem	0,35	0,34	0,35	0,34
<b>Seção: Planejamento e currículo (PC)</b>				
Ajustar itens de periodicidade	0,39	0,40	0,40	0,39
<b>Dimensão: Currículo, interações e práticas</b>				
Sem PC	0,48	0,47	0,47	0,47
Sem PARMA	0,52	0,51	0,51	0,51
Sem PC e sem PARMA	0,51	0,51	0,50	0,50
Ajuste* em PARMA	0,52	0,51	0,51	0,51
Ajuste** em PC	0,49	0,48	0,48	0,48
Ajuste em PARMA e em PC	0,52	0,51	0,51	0,51

\*Presença ou não de oportunidade de aprendizagem;

\*\*Itens de periodicidade;

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

Na seção de “Planejamento e currículo” foi imputado o valor zero para quando não há valores válidos. Dessa forma, é colocada uma categoria para quando não há periodicidade e teria o mesmo significado da opção “não” do item imediatamente anterior.

<sup>9</sup> As respostas referentes ao nível 1 receberam o número 1, e as respostas referentes aos níveis 2, 3 e 4 receberam o número 2.

<sup>10</sup> As respostas que foram nível 1 continua recebendo o número 1, nível 2 recebe o número 2, já nível 3 e 4 recebem o número 3.

Com isso, o indicador aproxima-se do limite superior do nível razoável de concordância, mas não o ultrapassa, como foi o caso anterior.

Por fim, foi reestimado o coeficiente Kappa para a dimensão “Currículo, interações e práticas” removendo as duas seções avaliadas e depois reinserindo-as após os ajustes testados. Feito isso, pode-se observar que a remoção da seção de PC impacta menos na concordância do que a de PARMA, quando se compara a ausência de ambos. E conseqüentemente, quando se ajusta somente em PC impacta menos do que somente em PARMA. Em conclusão, o nível de concordância entre os observadores para a dimensão alcança resultados semelhantes quando são removidas as seções PC e PARMA e quando são realizados os ajustes propostos anteriormente.

A Tabela 7 permite realizar análises a respeito da confiança “intraobservador” (onde um indivíduo aplica o instrumento no mesmo alvo mais de uma vez) e também sobre o formato de aplicação, e nisso está incluído a quantidade de aplicações em uma mesma turma e número de observadores. Considerando o número de visitas as turmas, pode-se observar que não há perdas no nível de concordância dos aplicadores, embora em alguns casos ocorram perdas marginais, como em “Alimentação” e “Segurança” para dois observadores e em todas as dimensões, exceto “Diversidade funcional”, para mais de dois observadores.

**Tabela 7** – Estimação do Kappa a nível de turma explorando o formato contínuo de aplicação.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
<i>Todos os formatos</i>					
2º dia	0,66	0,49	0,64	0,65	0,68
3º dia	0,65	0,48	0,65	0,63	0,67
4º dia	0,66	0,48	0,66	0,63	0,67
5º dia	0,66	0,48	0,65	0,63	0,66
<i>Um observador</i>					
2º dia	0,76	0,61	0,84	0,80	0,79
3º dia	0,81	0,61	0,85	0,81	0,81
4º dia	0,80	0,61	0,85	0,81	0,82
5º dia	0,81	0,61	0,86	0,82	0,81
<i>Dois observadores</i>					
2º dia	0,68	0,48	0,62	0,64	0,70
3º dia	0,66	0,47	0,62	0,61	0,66
4º dia	0,66	0,47	0,63	0,61	0,66
5º dia	0,66	0,47	0,63	0,61	0,66
<i>Mais de dois observadores</i>					
2º dia	0,55	0,36	0,46	0,49	0,56
3º dia	0,50	0,35	0,48	0,48	0,53
4º dia	0,52	0,35	0,48	0,47	0,52
5º dia	0,51	0,34	0,46	0,48	0,49

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

Em relação a quantidade de aplicadores, à medida que ela aumenta, diminui a concordância entre eles. Isso é um movimento esperado e pode ser ocasionado pelo efeito do acaso (quando  $P_{exp}$  na equação (7) aumenta) ou por discrepâncias nas avaliações (Fleiss, Nee e Landis, 1979). Quando é o mesmo indivíduo realizando todas as aplicações na turma, pode ocorrer dele se habituar ao ambiente e com isso aprimorar suas repostas referentes aos insumos. Por outro lado, pode ocorrer de o aplicador utilizar-se da memória ao invés da observação para preencher algum item, observadores diferentes podem corrigir esse possível problema, além de diversificar os julgamentos e percepções individuais.

Quanto aos níveis de concordância, no caso em que é o mesmo observador em todas as aplicações, as dimensões apresentam resultados excelentes, exceto a de “Currículos, interações e práticas” que foi nível bom de concordância. Quando é dois observadores, todas as dimensões se deslocam um nível abaixo, de excelente para bom, e de bom para moderado. Já para mais de dois observadores ocorre outra queda de nível de moderado para razoável na dimensão de “Currículo, interações e práticas”, e de bom para moderado nas demais.

Adicionar uma visita a mais quando se utiliza de mais de dois observadores pode reduzir marginalmente a concordância em até 0,07 pontos dentro do mesmo grupo e em até 0,17 pontos e um nível em relação a dois observadores. As dimensões possuem valores próximos do coeficiente Kappa, com exceção de “Currículo, interações e práticas”. Isso não significa que as avaliações desta dimensão não são confiáveis, mas que esses itens possuem alguma característica diferente dos demais que impactam na equivalência dos observadores nas respostas. Dito isso, no momento de traçar as estratégias de coleta dos dados utilizando a observação direta da EAPI é necessário considerar as perdas e ganhos com maior número de aplicadores.

### **3.2.2. Estabilidade**

O coeficiente de correlação intraclasse (ICC) foi estimado de forma semelhante ao Kappa (ver Tabela 17). Na Tabela 8, discriminado pelas dimensões normatizadas da EAPI, encontram-se os resultados da estabilidade, obtida pelo método de teste e reteste. De acordo com o Quadro 2, as dimensões de “Currículo, interações práticas” e “Segurança” obtiveram nível bom de correlação entre as aplicações ao longo do tempo, enquanto as demais atingiram o nível moderado de estabilidade. Contudo, os valores são mais próximo do limite superior (0,75) do que do inferior (0,51).

**Tabela 8** – ICC estimado a nível de turma.

	2º dia	3º dia	4º dia	5º dia
Alimentação	0,73	0,72	0,72	0,73
Currículo, interações e práticas	0,85	0,84	0,84	0,83
Diversidade funcional	0,74	0,74	0,73	0,74
Infraestrutura	0,70	0,69	0,70	0,70
Segurança	0,86	0,86	0,86	0,85

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

O desafio associado à análise de confiabilidade pelo método teste-reteste reside na possibilidade de surgirem efeitos de aprendizagem, transferência ou recordação, onde o primeiro teste pode influenciar o segundo, e assim por diante. Além disso, o intervalo de tempo entre as aplicações do instrumento também exerce impacto na confiabilidade. Um intervalo de tempo muito curto aumenta a probabilidade de ocorrerem efeitos de transferência relacionados à memória, prática ou humor, enquanto um intervalo mais longo amplia as chances de observar mudanças nas medidas (Marx *et al.*, 2003).

Para medidas que são relativamente estáveis ao longo do tempo, um intervalo mais curto entre testes e retestes pode ser apropriado. Griep *et al.* (2003) classificam a duração de duas semanas como nem muito curta e nem muito longa. Os instrumentos da EAPI buscam medir os insumos e as práticas. Esses são estáveis devido, respectivamente, aos investimentos e manutenções em materiais, equipamentos e estrutura predial e aos documentos curriculares que norteiam as atividades pedagógicas (por exemplo, o PPP que costuma ter validade de até dois anos).

A Tabela 9 mostra como foram os intervalos de tempo entre as aplicações do instrumento. As informações foram obtidas em relação ao primeiro dia de aplicação. A segunda aplicação ocorreu entre 1 a 8 dias após a primeira, com uma média de 1,5. As colunas de Q1 a Q4 dividem esse intervalo de tempo em quatro partes aproximadamente iguais e mostram quantas aplicações ocorreram para esse determinado período.

**Tabela 9** – Estatísticas descritivas das durações entre as aplicações da EAPI em cada turma.

	Média	Desv. Pad.	Mín	Máx	Q1	Q2	Q3	Q4
<b>2º dia</b>	1,5	1,0	1	8	132 (1 – 2)	8 (3 – 4)	3 (5 – 6)	1 (7 – 8)
<b>3º dia</b>	3,1	1,9	2	13	113 (1 – 3)	23 (4 – 6)	6 (7 – 9)	2 (10 – 13)
<b>4º dia</b>	5,7	3,2	3	16	79 (1 – 4)	31 (5 – 8)	27 (9 – 12)	7 (13 – 16)
<b>5º dia</b>	8,9	4,8	4	31	54 (1 – 7)	73 (8 – 14)	12 (15 – 21)	5 (22 – 31)
					432 (1 – 7)	122 (8 – 14)	17 (15 – 21)	5 (22 – 31)

Fonte: Elaborado a partir dos dados da EAPI Contínua.

Em relação às outras aplicações, seguiu uma crescente na amplitude do intervalo de tempo, com o máximo sendo 31. Em outras palavras, em até um mês as turmas receberam 5 aplicações da EAPI. A última linha da tabela mostra a distribuição de todas as aplicações com base no maior intervalo disponível. Disso, temos que a maioria das observações ocorreu em menos de 14 dias após a primeira, ou seja, abaixo da duração citada anteriormente e podendo ser considerado um curto período de tempo.

Tomando como base os intervalos contidos nas colunas Q1 a Q2 da Tabela 9, o ICC foi estimado para as dimensões. De acordo com a Tabela 10, quando se leva em consideração apenas a duração do tempo, para até 14 dias, obtêm-se correlações semelhantes às encontradas na Tabela 8. Depois desse período, as correlações se tornam menores, mas isso pode estar ligado ao tamanho da amostra (visto na última linha da Tabela 9).

**Tabela 10** – ICC estimado a nível de turma por quantidade de aplicações em cada intervalo de tempo.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
<i>Tudo</i>					
Q1	0,60	0,79	0,64	0,57	0,78
Q2	0,58	0,78	0,59	0,55	0,78
Q3	0,53	0,77	0,69	0,58	0,58
Q4	0,63	0,74	0,59	0,55	0,55
<i>2º dia</i>					
Q1	0,73	0,86	0,74	0,70	0,85
Q2	0,68	0,79	0,83	0,56	0,89
Q3	0,90	0,88	0,75	0,70	0,81
Q4	0,84	-	0,33	0,89	0,95
<i>3º dia</i>					
Q1	0,72	0,85	0,75	0,70	0,86
Q2	0,72	0,82	0,70	0,65	0,85
Q3	0,73	0,85	0,71	0,72	0,87
Q4	0,62	0,74	0,83	0,46	0,92
<i>4º dia</i>					
Q1	0,73	0,86	0,75	0,73	0,87
Q2	0,73	0,81	0,71	0,63	0,83
Q3	0,72	0,85	0,70	0,71	0,83
Q4	0,65	0,73	0,81	0,59	0,91
<i>5º dia</i>					
Q1	0,72	0,84	0,74	0,73	0,86
Q2	0,74	0,84	0,75	0,70	0,86
Q3	0,64	0,81	0,69	0,58	0,76
Q4	0,68	0,82	0,71	0,63	0,78

Fonte: Elaborado a partir dos dados da EAPI Contínua.

Quando, além da duração do tempo, também é levado em consideração quantas aplicações as turmas já receberam, observa-se um aumento marginal nos coeficientes de correlação para todas as dimensões. Os dois últimos intervalos de tempo, em geral, assumem valores que diferem mais dos dois primeiros, e novamente pode haver um impacto do tamanho da amostra. Contudo, a quarta e a quinta aplicações possuem uma distribuição mais suave nos intervalos e apresentam o mesmo comportamento.

Nenhum resultado mostrou estabilidade fraca e apenas “Segurança” apresentou grau excelente em três ocasiões, todas quando a aplicação ocorre no período mais distante da primeira. As demais foram predominantemente boas para “Currículo, interações e práticas” e “Segurança” e moderadas para o restante.

Por fim, a Tabela 11 mostra a estabilidade das aplicações, quando são considerados as quantidades de observadores que aplicaram a EAPI na turma. Os resultados apresentam um comportamento semelhante ao que ocorreu com o coeficiente

Kappa, que também já é esperado. À medida que aumenta o número de observadores, a correlação entre as aplicações diminui.

**Tabela 11** – ICC estimado a nível de turma por quantidade de aplicações e de observadores.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
<b>2º dia</b>					
1 obs.	0,79	0,88	0,87	0,84	0,90
2 obs.	0,77	0,86	0,71	0,68	0,86
> 2 obs.	0,61	0,80	0,63	0,57	0,81
<b>3º dia</b>					
1 obs.	0,84	0,87	0,87	0,85	0,91
2 obs.	0,73	0,86	0,70	0,66	0,85
> 2 obs.	0,58	0,79	0,64	0,55	0,81
<b>4º dia</b>					
1 obs.	0,83	0,87	0,86	0,85	0,91
2 obs.	0,73	0,85	0,70	0,66	0,85
> 2 obs.	0,60	0,79	0,64	0,57	0,80
<b>5º dia</b>					
1 obs.	0,84	0,87	0,87	0,85	0,91
2 obs.	0,73	0,84	0,70	0,66	0,85
> 2 obs.	0,60	0,78	0,63	0,57	0,78

Fonte: Elaborado a partir dos dados da EAPI Contínua.

Porém, nesse cenário, a dimensão “Currículo, interações e práticas” permanece no mesmo 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 dias. Menos observadores tornam as medidas do instrumento mais estáveis, porém há margem para que mais indivíduos possam aplicar o instrumento. Dessa forma, novamente devem ser ponderados por parte da coordenação de campo os ganhos e perdas com o formato pretendido.

### 3.2.3. *Consistência interna*

Para investigar se os instrumentos da EAPI estão medindo o que se propõem, foram estimados os coeficientes alfa de Cronbach para todas as dimensões. Na Tabela 12, encontra-se um método semelhante ao realizado para o Kappa da Tabela 2. Com base nas interpretações (ver Quadro 2), percebe-se que os itens medem as dimensões no mesmo nível de consistência, independente das segmentações propostas, exceto para “Diversidade funcional” que reduz. O melhor desempenho ocorre para as dimensões de

“Equipe e gestão” com consistência excelente, seguido por “Currículo, interações e práticas” e “infraestrutura” com consistência interna boa para a maioria das aplicações. E logo após, “Alimentação” e “Segurança” com nível considerado fraco.

**Tabela 12** – Alfa de Cronbach a nível de rede por aplicação e faixa etária.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Equipe e gestão	Infraestrutura	Segurança
<i>Toda a amostra</i>						
1º dia	0,40	0,67	0,76	0,80	0,75	0,53
2º dia	0,47	0,72	0,75	0,79	0,77	0,51
3º dia	0,41	0,78	0,74	0,79	0,76	0,50
4º dia	0,46	0,68	0,74	0,80	0,78	0,50
5º dia	0,45	0,76	0,73	0,80	0,78	0,45
<i>Creche</i>						
1º dia	0,42	0,77	0,63	0,83	0,77	0,45
2º dia	0,50	0,77	0,61	0,84	0,78	0,46
3º dia	0,45	0,76	0,60	0,83	0,80	0,51
4º dia	0,48	0,70	0,62	0,83	0,77	0,53
5º dia	0,52	0,78	0,66	0,84	0,76	0,55
<i>Pré-escola</i>						
1º dia	0,46	0,72	0,46	0,82	0,71	0,42
2º dia	0,53	0,78	0,55	0,82	0,67	0,56
3º dia	0,48	0,80	0,46	0,82	0,69	0,46
4º dia	0,41	0,73	0,46	0,82	0,72	0,52
5º dia	0,42	0,77	0,54	0,82	0,71	0,42

Fonte: Elaborado a partir dos dados da EAPI Contínua.

Porém, a organização dos dados nessa configuração não favorece uma estimação adequada do alfa de Cronbach, visto que ele é elaborado para escalas onde os itens são as unidades de medida, e para obter resultados robustos, deve ser maior do que o número de avaliadores. Quando isso ocorre, o alfa pode não fornecer uma estimativa precisa da consistência interna da escala e também pode ser sensível a respostas extremas de um pequeno grupo de aplicadores, o que faz sua interpretação ser difícil e não representativa (DeVellis, 2021; Nunnally e Bernstein, 1994).

Dessa forma, foi realizado a estimação do alfa para cada turma. Diferente das medidas anteriores, o coeficiente alfa da dimensão a nível de turma não é a média dos alfas de cada turma. Sendo assim, na Tabela 13, encontram-se as descritivas dos coeficientes estimados. Pode-se observar que 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, assumindo nível considerado perfeito ou redundante.

Quando isso ocorre, significa que os itens estão oferecendo informações muito semelhantes ou quase idênticas sobre o constructo que está sendo medido. Isso pode resultar em uma avaliação inflada da consistência interna por conta de repetição

excessiva, o que pode indicar falta de diversidade no conjunto de itens. Idealmente, cada item em um instrumento deve contribuir com uma informação única para a avaliação do construto em questão. Nesse caso, é possível realizar um teste, no qual itens são removidos e o coeficiente é reestimado.

**Tabela 13** – Alfa de Cronbach a nível de turma (n=144).

	Média	Desv.Pad.	Min	Max	Qtd de turmas					
					(1) <sup>11</sup>	(2)	(3)	(4)	(5)	(6)
<b>Alimentação</b>	0,74	0,18	0,21	0,98	8	16	21	34	50	12
<b>Currículo, interações e práticas</b>	0,93	0,03	0,83	0,97	0	0	0	0	17	127
<b>Diversidade funcional</b>	0,75	0,17	0,18	0,99	5	16	14	20	61	22
<b>Infraestrutura</b>	0,90	0,07	0,65	0,99	0	0	5	8	45	86
<b>Segurança</b>	0,72	0,17	0,16	0,94	6	17	17	50	35	12

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

As demais dimensões apresentam boa consistência interna em média, com pelo menos 70% das turmas nesse nível ou acima e menos de 15% em níveis fracos ou insignificantes. Esses resultados se aproximam do que é mostrado por Dockrell *et al.* (2023) para o MELE. Em seguida, será mostrado o coeficiente alfa considerando a amostra para quantidade de aplicações e de observadores. Não foi possível estimar para a combinação do primeiro e do segundo dia, por conta de erros na matriz de correlação dos itens devido a variação total dos itens ser zero, isso ocasionou coeficientes iguais a 1 para todas as dimensões. Sendo assim, os resultados não são confiáveis.

Porém, a partir da inclusão do terceiro dia, há resultados com intervalo de confiança de 95%. Isso mostra um possível número mínimo de aplicações necessários para utilização desse método de avaliação para a quantidade itens das dimensões da EAPI (quantidade de linhas da matriz na equação (3)). De acordo com a Tabela 14, a consistência interna das dimensões medidas pela EAPI melhora quanto mais diversificados forem os observadores em reaplicações.

<sup>11</sup> Referente ao Quadro 2: (1) – Insignificante, (2) – Fraco, (3) – Moderado, (4) – Bom, (5) – Excelente, (6) – Perfeito ou redundante.

**Tabela 14** – Alfa estimado a nível de turma por quantidade de aplicações e de observadores.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
<b>3º dia</b>					
1 obs.	0,84	0,95	0,68	0,88	0,86
2 obs.	0,70	0,97	0,92	0,97	0,83
> 2 obs.	0,86	0,97	0,88	0,96	0,84
<b>4º dia</b>					
1 obs.	0,71	0,92	0,66	0,83	0,66
2 obs.	0,82	0,96	0,88	0,95	0,77
> 2 obs.	0,80	0,95	0,84	0,93	0,76
<b>5º dia</b>					
1 obs.	0,62	0,90	0,57	0,83	0,67
2 obs.	0,80	0,94	0,84	0,94	0,79
> 2 obs.	0,76	0,93	0,82	0,92	0,72

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

O número de reaplicações não altera o nível de consistência das dimensões, exceto para “Alimentação” e “Segurança”. Na primeira, diminui um nível para cada aplicação adicional pelo mesmo observador, aumenta um nível do 3º para o 4º dia e depois estabiliza ao ter dois observadores, e diminui um nível no 5º dia adicional para mais de dois aplicadores. Na segunda, o nível de consistência é reduzido após a quarta aplicação e se mantém depois da quinta. Essa redução é maior quando é o mesmo observador.

Por fim, a Tabela 15 mostra o que acontece com o alfa de Cronbach se determinados itens forem removidos da construção das dimensões. Foram realizadas estimativas para todas as combinações de turmas e itens e calculada a variação do alfa, depois foi feita a média dessas variações. Apenas dois itens implicaram em uma variação marginal positiva no alfa caso fossem removidos nas dimensões de “Segurança” e “Infraestrutura”. Isso significa que caso não fizessem parte desses constructos, a consistência interna aumentaria, mesmo que apenas na segunda e terceira casa decimal, respectivamente.

**Tabela 15** – Alfa de Cronbach reestimado removendo um item.

Dimensão	Item	Varição média no alfa
<i>Itens que se removidos, aumentam o alfa</i>		
Segurança	Calhas e “canaletas” para escoamento da chuva obstruídas;	0,015
Infraestrutura	Objetos empacotados / mal armazenados na escola;	0,001
<i>Itens que se removidos, reduzem o alfa</i>		
Alimentação	Cardápio compreensível pelas crianças;	-0,100
	Guardanapos acessíveis para uso das crianças;	-0,100
	Lixeira com tampa e pedal na altura das crianças;	-0,082
Currículo, interações e práticas	Estratégias verbais de gestão dos conflitos e dos comportamentos das crianças pela(o) professora(or);	-0,004
	Estratégias de acolhimento afetivo às crianças;	-0,004
	Crianças podem beber água quando desejam, mesmo que precisem pedir para um adulto;	-0,003
Diversidade funcional	Piso tátil direcional nas dependências da UE em pátios e espaços amplos;	-0,086
	Rampas como alternativa para degraus e escadas;	-0,064
	Elevador ou plataformas elevatórias entre andares;	-0,044
Infraestrutura	Há área externa descoberta que possibilite a brincadeira das crianças;	-0,050
	Há área externa coberta que possibilite a Brincadeira das crianças;	-0,016
	Roupa de cama limpa e de uso exclusivo da criança;	-0,010
Segurança	Área de circulação de veículos sem isolamento por grades ou muros;	-0,080
	Extintores de incêndio vencidos;	-0,075
	Fiação exposta ou mal conservada;	-0,068

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

Como exercício, foram explorados os três itens que são mais indispensáveis, em relação a consistência interna. Ou seja, itens que se removidos da construção das dimensões, tornariam o alfa menor. Dos mostrados na Tabela 15, aqueles com maior perda substancial de consistência são os que medem se o cardápio é compreensivo pelas crianças, se há guardanapo, se há piso tátil direcional, se tem lixeira no refeitório, se há isolamento de áreas com circulação de veículos e se os extintores estão vencidos.

No geral, os constructos medidos pela EAPI demonstraram consistência interna boa ou excelente, com leve redundância<sup>12</sup> em alguns casos. Em reaplicações, o ideal é que seja diversificado o observador para pelo menos dois indivíduos diferentes. E a partir do quarto dia, a consistência interna costuma estabilizar-se.

<sup>12</sup> Mais próximo do limite inferior do intervalo.

### 3.3. Considerações finais

Um aspecto importante da aplicação de protocolos de observação em salas de aula é a qualidade dos dados colhidos (Volpe *et al.*, 2005). A EAPI foi construída a partir de uma iniciativa anterior de sucesso, o MELQO, com o intuito de mensurar medidas de qualidade na primeira infância. Esse ensaio buscou analisar a confiabilidade da aplicação contínua do instrumento de observação direta da EAPI. E para isso foram utilizadas a equivalência medida pelo coeficiente Kappa, a estabilidade pelo ICC e a consistência interna pelo alfa de Cronbach. A amostra foi composta por 720 observações realizadas em 144 turmas, 5 vezes em cada.

Os principais resultados mostraram que para a maioria das dimensões<sup>13</sup>, o nível de concordância foi considerado bom<sup>14</sup> e novas aplicações na turma parece não alterar isso de forma significativa. A dimensão de “Currículo, interações e práticas” contém itens que trazem desafios a análise, entretanto é possível contorná-los abordando as seções que a compõem. Ao fazer isso, pode se concluir que o nível de concordância é bom, no caso em que seja considerado apenas se os aplicadores estão identificando ou não os momentos de oportunidade aprendizagem. O mesmo não se pode concluir a respeito do julgamento da qualidade dessa oportunidade.

As aplicações se mantiveram estáveis entre o mesmo observador e entre dois aplicadores diferentes. Mais do que isso resultou em reduções do nível de estabilidade para as dimensões de “Alimentação”, “Diversidade funcional” e “Infraestrutura”. A consistência interna foi reduzida em todas as dimensões pelo aumento do número de aplicações contínuas, por outro lado, a diversificação dos aplicadores a aumentou e indicou estabilizar-se depois do quarto dia.

Dessa forma, quando se trata da confiabilidade, este estudo sugere que é possível expandir o formato de aplicação da EAPI. De fato, as variações nas respostas dos observadores mantiveram-se dentro dos limites adequados estabelecidos pela literatura (ver Quadro 2). Explorar o uso de duas ou três pessoas diferentes para a aplicação na mesma turma não compromete a concordância e melhora a consistência interna do instrumento. Mas, como o acréscimo de uma observação para medir a qualidade dos ambientes em uma turma de referência pode reduzir o nível de confiabilidade, deve-se ponderar essa ação no momento de determinar o formato de coleta.

---

<sup>13</sup> ver Quadro 1

<sup>14</sup> ver Quadro 2

Em particular, a dimensão “Currículo, interações e práticas” aborda a interação entre professora e criança, ou seja, é um evento que depende da dinâmica que ocorre em uma data específica. Os efeitos causados pela rotina podem não ser capturados por essa metodologia e causaram redundância aos valores estimados. Uma sugestão para mitigar esses problemas e abrir margem para outros estudos é considerar a presença de dois ou mais observadores para avaliar a turma.

## 4. ENSAIO II – UMA ANÁLISE DO EFEITO DAS CARACTERÍSTICAS DOS OBSERVADORES NAS MEDIDAS DE CONFIABILIDADE DA EAPI CONTÍNUA

### 4.1. Introdução

A avaliação dos ambientes de aprendizagem desempenha um papel crucial no desenvolvimento e aprimoramento da educação infantil (Yoshikawa *et al.*, 2018). Instrumentos de medição são ferramentas fundamentais nesse processo. Contudo, a eficácia desses instrumentos não é apenas determinada pela sua qualidade intrínseca, mas também pela capacidade e condições dos aplicadores responsáveis por sua implementação (Vitiello *et al.*, 2018).

Durante a aplicação do instrumento de observação direta da EAPI, uma pessoa participa das atividades de uma turma de referência por três horas e meia, observando e realizando anotações sobre o que está acontecendo. Em seguida, de acordo com a orientação do treinamento, o aplicador preenche o instrumento com as informações colhidas. Alguns vieses podem influenciar essas respostas.

O viés de expectativas, também conhecido como viés de observador, ocorre quando os resultados de um estudo são influenciados pelas expectativas prévias ou pelas crenças dos participantes (Harris; Rosenthal, 1985). Devido à tendência natural das pessoas de ajustar seu comportamento quando estão cientes da observação, o comportamento dos professores e das crianças pode ser influenciado pela presença do aplicador. Como resultado, as interações e dinâmicas na sala de aula podem ser afetadas, ocasionando em situações que não representam fielmente a realidade cotidiana da sala e distorcendo as percepções do observador (O’leary, 2020).

A ocorrência de variabilidades individuais dos aplicadores, ou seja, diferenças nos padrões de observação e avaliação entre diferentes indivíduos podem surgir devido a diferenças no julgamento, experiência, formação e tendências pessoais dos observadores. Essas variabilidades individuais afetam a consistência e confiabilidade das avaliações, o que por sua vez, compromete a validade dos resultados (Downer *et al.*, 2010).

Em estudos observacionais, as interações entre professor e criança registradas por observadores com treinamento específico tendem a produzir avaliações mais consistentes e precisas (Pianta; La Paro; Hamre, 2008). Além disso, a experiência prévia na área da educação infantil e o conhecimento sobre o desenvolvimento infantil também podem

melhorar a capacidade de observadores avaliarem os ambientes de aprendizagem de forma sensível e abrangente (LoCasale-Crouch *et al.*, 2008).

A fadiga física e mental está associada com dificuldades em tomar decisões e realizar avaliações (Moore; Gupta; Neto, 2013). Em situações que os avaliadores estão exaustos ou sobrecarregados mentalmente, suas habilidades de foco, atenção e discernimento podem ser afetadas (Sweller, 1994).

Neste ensaio, em formato de *letter*, serão empregadas informações sobre os aplicadores da EAPI contínua e o processo de coleta, visando estabelecer associações com os coeficientes de confiabilidade investigados no ensaio I. As próximas subseções apresentam com mais detalhes a metodologia utilizada e os resultados encontrados, respectivamente.

## **4.2. Metodologia**

### **4.2.1. Base de dados**

Serão utilizados os coeficientes de confiabilidade abordados no ensaio I estimados por turma e por dia adicionado. A distribuição dos coeficientes Kappa, ICC e alfa de Cronbach podem ser vistos nos Gráficos 2, 3 e 4 presentes no apêndice B. Elas estão representadas pelos intervalos definidos pelo Quadro 2 e possuem desvio padrão de 0,20; 0,17 e 0,16 respectivamente.

As demais variáveis presentes na base de dados buscam capturar o efeito dos fatores que podem influenciar a consistência e a precisão das respostas dos observadores no momento de preencher o instrumento da EAPI. Para os vieses de resposta, variabilidades individuais, experiência, treinamento e carga cognitiva foram utilizadas o sexo dos observadores, o conhecimento prévio sobre desenvolvimento infantil por meio do Kidi, a formação no qual o indivíduo participou e a nota na primeira tentativa de certificação. E para medir efeitos de fadiga foi utilizada a distância entre a residência do observador até a unidade educacional e quantidade de observações realizadas em cada semana.

### **Inventário do Conhecimento do Desenvolvimento Infantil (Kidi)**

O instrumento contém 75 questões que envolvem quatro categorias: práticas de cuidado; saúde e segurança; normas e aquisições; e princípios do desenvolvimento. Neste questionário são feitas perguntas sobre o desenvolvimento infantil normal e sobre cuidados em relação à saúde das crianças.

Cada item descreve o que poderia ser o comportamento de uma criança comum ou o que poderia afetar o crescimento e o comportamento de um bebê. O respondente deve marcar em cada item se concorda com a afirmação, se discorda dela, ou se não está seguro da resposta. É orientado que responda cada pergunta baseando-se em seu conhecimento sobre as crianças de uma maneira geral (MacPhee, 1981).

### **Formação e treinamento**

A EAPI foi desenvolvida com a intenção de ser o mais objetiva e intuitiva possível. No entanto, é crucial que exista um processo de aprendizado e adaptação antes de alcançar plena aptidão para sua utilização. Nessa perspectiva, os observadores necessitam de treinamento para compreender a escala de maneira abrangente e estarem capacitados para sua aplicação. O primeiro contato com a EAPI sempre se dá por meio do treinamento para ser observador, após adquirir experiência em campo é que há a possibilidade de se tornar formador.

A capacitação para observador tem duração de cinco dias em ambos os turnos. Nos primeiros dias são apresentados os instrumentos que compõem a EAPI, os conceitos-chave, definições e orientações acerca da condução ética da aplicação. Ao final do período, os participantes são enviados para um treino em campo, e depois comparam suas respostas entre si em busca de um consenso. Por fim, há uma prova de certificação que abrange todo o conteúdo abordado durante a formação.

### **Estatísticas descritivas**

De acordo com a Tabela 16, observa-se que houve um maior número de aplicações realizadas por pessoas do sexo feminino (445), em comparação com o sexo masculino (275). No entanto, quando é analisada a média de observações, percebe-se que os homens tiveram uma média de 30,6 observações, enquanto as mulheres tiveram uma média de 21,4 observações. Essa discrepância pode ser atribuída ao fato de que havia mais aplicadoras (21) do que aplicadores (9) envolvidos no processo.

**Tabela 16** – Estatísticas descritivas dos dados utilizados.

	N	Média	Desv.Pad.	Min	Max
<b>Painel A: Características de campo</b>					
Quantidade observações					
<i>Masculino</i>	275	30,6		10	47
<i>Feminino</i>	445	21,2		1	48
Distância de casa até a UE	-	5.274,8		273,1	15.650,8
<b>Painel B: Características dos observadores</b>					
Quantidade de observadores					
<i>Masculino</i>	9	-		-	-
<i>Feminino</i>	21	-		-	-
Kidi	19	58%	11,8%	36%	76%
Formação					
1 <sup>a</sup>	10	-	-	-	-
2 <sup>a</sup>	4	-	-	-	-
3 <sup>a</sup>	16	-	-	-	-
Nota certificação	30	6,9	1,1	5,0	9,8

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

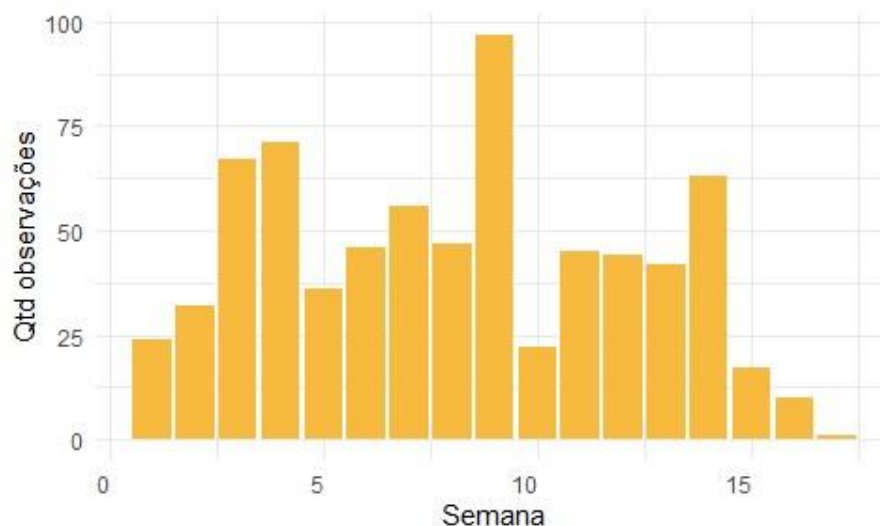
Dos 30 observadores, 19 preencheram o questionário do Kidi. O desempenho médio foi de 58% de acertos, o mínimo 36% e o máximo 76%, com desvio padrão de 11,8%. Os 30 observadores participaram de três formações diferentes, na primeira foram 10, na segunda 4 e na terceira 16. Esses grupos também determinam a experiência em campo, visto que eles estão dispostos em ordem cronológica e os primeiros participaram da coleta de campo do estudo nacional.

A certificação da EAPI é obtida por meio de uma prova, na qual são postas situações de campo e deve-se analisar a pontuação de itens referentes ao momento que está sendo exposto. O candidato tem três tentativas para conseguir a nota maior ou igual a 7,0 e se certificar. Ao longo das tentativas, a dificuldade pode variar, portanto foi utilizada aqui apenas a primeira tentativa. Por isso, a nota mínima apresenta é 5 e o máximo é 9,8 e a média foi 6,9.

Em relação ao deslocamento dos observadores, foram percorridos aproximadamente 5,3 km em média das residências até as unidades educacionais, com mínimo de 273,1 metros e máximo de 15,6 km. O Gráfico 1 mostra a distribuição do quantitativo de observações ao longo das semanas de coleta de campo. O máximo ocorreu na 9<sup>a</sup> semana, que teve 97 observações, e o mínimo na última semana, na qual teve apenas

uma aplicação. Em média, foram 42,4 aplicações por semana (44,9 sem considerar a última semana que se trata de um outlier).

**Gráfico 1** – Quantidades de observações realizadas em cada semana.



Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

A Tabela 17 contém as correlações entre a formação, taxa de acerto no Kidi e nota na certificação. Os valores apresentados são baixos e no teste de significância resultou que não são estatisticamente diferentes de zero. Isso significa que não há uma relação linear entre elas.

**Tabela 17** – Correlação entre as características dos observadores.

	Kidi	Formação	Certificação
Kidi	1,00		
Formação	-0,32	1,00	
Certificação	0,21	-0,18	1,00

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

#### 4.2.2. Procedimentos estatísticos

A amostra consiste em dados longitudinais coletados em cinco momentos ao longo de um período de tempo. A estimação do modelo será realizada utilizando o método de *pooling*, que assume homogeneidade nos coeficientes para todas as unidades individuais. Os resultados estimados serão apresentados interpretando os coeficientes relevantes e destacando a significância estatística. A especificação do modelo é representada por:

$$Y_{it} = \alpha + \beta X_{it} + \mu_i + \delta_t + \varepsilon \quad (7)$$

Onde  $Y_{it}$  são os coeficientes de confiabilidade (Kappa, ICC e alfa de Cronbach),  $X_{it}$  é um vetor com as variáveis explicativas.  $\mu_i$  e  $\delta_t$  são os efeitos fixos individuais e temporais, respectivamente.  $\varepsilon$  é o termo de erro.

$X_{it}$  é composto pela quantidade de reaplicações, quantidade de observadores, identificação se é sexo feminino, distância em metros da casa do observador até a unidade, quantidade de aplicações na semana corrente, nota (taxa de acerto) no Kidi, identificação da formação na qual o observador participou e nota na primeira tentativa da certificação.

Há efeitos fixos para cada turma observada, no intuito de capturar as especificidades individuais de cada turma e professora e também para o dia da semana, nesse caso, identificando se o dia é segunda, terça, etc., de maneira a se adequar as características de rotina adotadas por cada escola, efeitos fixos para o aplicador, e por fim, efeitos fixos temporais para cada reaplicação.

### 4.3. Resultados

O modelo foi estimado para cada dimensão com a intenção de utilizar uma medida já validada. De acordo com a Tabela 18, que possui os valores dos coeficientes para o indicador de confiabilidade Kappa, a quantidade de observadores, a distância, a nota no Kidi, experiência, nota da certificação e número de observações por semana foram significantes em algumas dimensões. Ou seja, a concordância “interobservador” sofre um impacto estatisticamente significativo dessas variáveis. 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”, respectivamente.

**Tabela 18** – Resultados da estimação para o coeficiente Kappa.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
Qtd. observações	0,001 (0,005)	0,001 (0,003)	-0,002 (0,008)	-0,005 (0,003)	0,001 (0,006)
Qtd. observadores	0,016 (0,027)	-0,034* (0,019)	0,049* (0,026)	-0,014 (0,021)	-0,141*** (0,032)
Feminino	-0,027 (0,033)	-0,001 (0,023)	0,011 (0,031)	0,019 (0,025)	0,018 (0,039)
Distância casa/UE	0,000** (0,000)	0,000 (0,000)	0,000* (0,000)	0,000 (0,000)	0,000 (0,000)
Nota Kidi	0,263* (0,150)	-0,029 (0,104)	0,028 (0,143)	0,097 (0,114)	0,072 (0,176)
Experiência	0,032 (0,035)	-0,017 (0,025)	-0,041 (0,044)	0,018 (0,027)	0,078* (0,042)
Nota certificação	-0,009 (0,020)	-0,016 (0,014)	-0,013 (0,019)	0,010 (0,015)	0,044* (0,024)
Qtd. obs. por semana	-0,012* (0,020)	-0,004 (0,004)	-0,006 (0,006)	-0,006 (0,005)	0,002 (0,007)
R <sup>2</sup> ajustado	0,811	0,824	0,863	0,860	0,779
N	144	144	144	144	144

Nível de significância: \*\*\* p<0,01; \*\* p<0,05; \* p<0,10.

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

A quantidade de observações não se mostrou estatisticamente significativa. Esse resultado pode ser associado ao que foi visto no ensaio, onde o Kappa se mostrou pouco sensível às reaplicações. A distância percorrida não teve impacto, por outro lado, a experiência e a quantidade de aplicações por semana apresentaram, respectivamente, efeito de 0,08 e 0,04 na dimensão de “Segurança”.

A nota no Kidi e a quantidade de aplicações por semana tiveram, respectivamente, efeito de 0,26 e -0,01 no nível de concordância para a dimensão de “Alimentação”. Isso significa que indivíduos com maior conhecimento prévio sobre desenvolvimento infantil tendem a concordar em suas respostas referentes a itens que medem esse constructo, como práticas de lavagem de mãos, oportunidades de aprendizagem que envolvam momentos de alimentação, identificação da presença de cardápio e se é compreensivo pelas crianças.

Os resultados do modelo para o coeficiente que analisa a estabilidade estão na Tabela 19. Cada coluna representa o ICC da turma para determinada dimensão. “Currículo, interações e práticas” obteve mais variáveis significativas, são elas: quantidade de observações, experiência do aplicador, e nota da certificação.

Essa dimensão contém os itens de oportunidades de aprendizagem com base no brincar, e tanto a certificação como o nível de experiência são atributos do observador que impactam na estabilidade do julgamento em reaplicações da EAPI, 0,06 e 0,03, respectivamente. Ou seja, quem participou de aplicações anteriormente, de fato aprimora a compreensão das definições abordadas nessa dimensão, em comparação com quem está em sua primeira coleta de campo.

**Tabela 19** – Resultados da estimação para o ICC.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
Qtd observações	0,007 (0,004)	-0,010** (0,004)	-0,007 (0,005)	0,000 (0,004)	0,001 (0,003)
Qtd observadores	-0,009 (0,024)	-0,002 (0,022)	0,052* (0,027)	-0,019 (0,022)	-0,025 (0,020)
Feminino	-0,024 (0,029)	0,009 (0,030)	0,001 (0,033)	0,021 (0,027)	-0,023 (0,025)
Distância	0,000** (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
Kidi	0,010 (0,132)	-0,128 (0,128)	-0,106 (0,150)	-0,196 (0,123)	-0,026 (0,112)
Experiência	0,041 (0,031)	0,057* (0,031)	-0,057 (0,036)	-0,022 (0,029)	0,021 (0,027)
Certificação	0,011 (0,018)	0,033* (0,017)	-0,021 (0,020)	0,000 (0,017)	0,013 (0,015)
Qtd. obs. semana	-0,011** (0,006)	0,000 (0,006)	-0,003 (0,006)	-0,004 (0,005)	0,003 (0,004)
R <sup>2</sup> ajustado	0,828	0,741	0,807	0,841	0,768
N	144	144	144	144	144

\*\*\* p<0,01; \*\* p<0,05; \* p<0,10.

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

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. A distância percorrida pelo observador não pareceu ter efeito sobre a estabilidade. Por outro lado, o número de observações que os aplicadores realizaram por semana reduziu em 0,01 o ICC para “Alimentação”. Esse constructo contém itens que envolvem interpretação de práticas e oportunidades de aprendizagem pelos observadores.

Para concluir, o coeficiente Alfa de Cronbach do instrumento sofreu efeito, estatisticamente significativo, da quantidade de observações em todas as dimensões (Tabela 20). Em suma, uma aplicação contínua 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.

**Tabela 20** – Resultados da estimação para o Alfa de Cronbach.

	Alimentação	Currículo, interações e práticas	Diversidade funcional	Infraestrutura	Segurança
Qtd observações	-0,050*** (0,010)	-0,018*** (0,001)	-0,064** (0,011)	-0,023*** (0,005)	-0,062*** (0,012)
Qtd observadores	-0,062 (0,041)	-0,003 (0,004)	0,003 (0,042)	-0,002 (0,019)	0,006 (0,046)
Feminino	0,065 (0,060)	0,007 (0,007)	-0,022 (0,062)	0,002 (0,028)	-0,005 (0,068)
Distância	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
Kidi	-0,390 (0,256)	-0,007 (0,028)	-0,173 (0,261)	0,014 (0,120)	-0,061 (0,290)
Experiência	-0,080 (0,060)	-0,001 (0,006)	0,012 (0,061)	-0,017 (0,028)	-0,166** (0,068)
Certificação	-0,001 (0,034)	0,002 (0,004)	0,001 (0,036)	-0,008 (0,016)	-0,083** (0,039)
Qtd. obs. semana	0,010 (0,010)	0,000 (0,001)	-0,004 (0,011)	-0,001 (0,005)	0,018 (0,012)
R <sup>2</sup> ajustado	0,599	0,786	0,659	0,635	0,576
N	139	139	139	139	139

\*\*\* p<0,01; \*\* p<0,05; \* p<0,10.

Fonte: Elaborado pelo autor a partir dos dados da EAPI Contínua.

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. Esse resultado pode estar ligado também com a formação que os aplicadores receberam, a orientação do facilitador pode influenciar na forma como os itens são interpretados. Dado que a EAPI se propõe a ter uma formação padronizada, abrem-se pretextos para um estudo mais aprofundado desse processo.

#### 4.4. Considerações finais

À medida que o atendimento na educação infantil se expandiu no Brasil na última década, em grande parte por conta da Meta 1 do PNE que buscou universalizar o ensino infantil na pré-escola e atender no mínimo 50% das crianças com até 3 anos e 11 meses, a preocupação que surge é em relação a qualidade dos ambientes de aprendizagem que são ofertados.

Obter essas medidas de qualidade conta com a colaboração de pessoas que realizem a coleta de informações (Vitiello *et al.*, 2018). E isso pode ser feito por meio de instrumentos elaborados de forma rigorosa e de acordo com os parâmetros educacionais vigentes e que possuam confiabilidade e validade. Neste estudo foi analisado o efeito de características inerente aos observadores da EAPI nos coeficientes de confiança obtidos a partir dos constructos medidos pela escala.

A quantidade de aplicações impactou negativamente a consistência interno do instrumento e o número de observadores teve efeito sobre a concordância. Isso corrobora com o que foi encontrado no ensaio I. O sexo dos indivíduos não obteve resultados estatisticamente significantes e a distância parece não influenciar a confiabilidade dos dados.

O conhecimento sobre desenvolvimento infantil, a experiência e o desempenho na certificação contribuíram positivamente para confiabilidade e a carga de aplicações semanais negativamente. Ou seja, a afinidade com a área é um pré-requisito que enriquece a qualidade dos dados obtidos, ao passo que o volume de trabalho pode comprometê-los.

O fato de as variáveis ligadas ao processo formativo dos observadores terem obtido resultados estatisticamente significantes para explicar a confiabilidade, demonstra a importância do treinamento para se adequar ao uso dos instrumentos da EAPI. Isso cria oportunidades de analisar com mais detalhes as fases de certificação e qualificação dos aplicadores afim de aprimorá-las e contribuir para o desenvolvimento profissional.

#### 5. LIMITAÇÕES

Este estudo busca contribuir para aprimoramento da aplicação do instrumento de observação direta da EAPI em turmas de educação infantil. Os resultados iniciais fornecem evidências de que a EAPI é confiável para medir qualidade dos ambientes de aprendizagem. Contudo, é importante reconhecer algumas limitações que não foram ou não puderam ser abordadas anteriormente.

A amostra utilizada consiste em dados referentes a um único município, que pode possuir características homogêneas. Embora a EAPI tenha sido elaborada para ser utilizada em todo território brasileiro, não é possível concluir que os aplicadores em todas as regiões do país possuam as mesmas características. Além disso, podem haver características regionais não observáveis das unidades educacionais, isso inclui ainda áreas urbanas e rurais.

Seria ideal para estudos futuros, uma amostra de tamanho maior e que abrangesse um conjunto mais diversificado de regiões geográficas. Além disso, os fatores de viés de resposta e variabilidade individual podem não ter sido totalmente capturadas pelos efeitos fixos e influenciar a confiabilidade de outras formas. Como cada aplicador estava observando situações diferentes, não foi possível isolar esse efeito por completo. Principalmente quando se trata da dimensão de “Currículo, interações e práticas”.

Downer *et al.* (2010) utilizaram uma estratégia que consistia em dois aplicadores observarem e avaliaram de forma independente a mesma situação. Em estudos futuros, essa metodologia pode ser aplicada a pelo menos uma parte da amostra (os autores citados realizaram em 20%) e assim obter medidas dos ambientes e especialmente das interações entre professora e criança que possam ser diretamente comparadas.

É interessante ainda para estudos futuros com a EAPI contínua, elaborar estratégias que permitam analisar o que é casual e o que excepcional na rotina da turma de referência. Ou seja, identificar oportunidades de aprendizagem e interações que ocorrem normalmente nos ambientes da unidade e se a professora possui influências nelas. Para isso, é necessário isolar a presença do observador e os possíveis vieses e contar com uma amostra de tamanho maior da rotina da turma de referência.

## 6. CONCLUSÃO

Em um cenário de protagonismo das ações voltadas para o aprimoramento da qualidade na educação infantil, formadores de políticas públicas precisam contar com informações precisas e confiáveis a respeito do sistema educacional. Na literatura, têm se discutido sobre instrumentos capazes de mensurar medidas de qualidade (Shirasu *et al.*, 2021). Para garantir que o acesso à educação infantil resulte em desenvolvimento e das crianças, os ambientes de aprendizagem e as interações com as professoras devem fornecer segurança, ser estimulantes, possuir estrutura física adequada, materiais pedagógicos diversificados e interações positivas (Maia; Williams, 2005).

Os ensaios desenvolvidos nesse trabalho exploraram a confiabilidade da EAPI, com análise de estabilidade, equivalência e consistência interna. Além disso, foram utilizadas características dos observadores, como conhecimento sobre desenvolvimento infantil, desempenho na certificação da EAPI, experiência e o volume de observações semanais, para estimar os coeficientes de confiança do instrumento.

Os resultados evidenciaram que o formato contínuo manteve a confiabilidade da EAPI para a maior parte dos cenários avaliados. Foi encontrado que o número de reaplicações não afeta a concordância entre os observadores e nem a estabilidade, por outro lado, o número de aplicadores influencia a confiabilidade em todas as medidas. A consistência interna sofreu efeito negativo de aplicações contínuas, mas indicou estabilizar-se após o quarto dia de observação.

Dado que, em reaplicações, dois ou três observadores obtiveram melhores coeficientes e a quantidade de observações adicionais reduziu esses números. O ideal é ponderar esses dois resultados em um formato que varie de duas a quatro aplicações. Contudo, outra sugestão é combinar formatações distintas à diferentes partes da amostra. Selecionar um grupo para receber mais observações, outro para receber menos, um para receber dois observadores simultaneamente e outro para receber apenas um por vez.

O presente trabalho descreveu uma adaptação brasileira, a EAPI, de uma escala validada em contexto global, o MELQO, para medir qualidade na educação infantil. E contribui com a literatura agregando evidências da confiabilidade do instrumento e apresentando perfis de observadores que melhoram a qualidade dos dados coletados.

## REFERÊNCIAS

- ALEXANDRE, Neusa Maria Costa *et al.* A confiabilidade no desenvolvimento e avaliação de instrumentos de medida na área da saúde. **Revista Eletrônica de Enfermagem**, v. 15, n. 3, p. 800-7, 2013.
- BRASIL. Ministério da Educação. Base Nacional Comum Curricular. Brasília, 2018.
- CAMPOS, Maria Malta. Avaliação da qualidade na educação infantil: impasses e perspectivas no Brasil. **Pesquisa e debate em educação**, v. 10, n. 1, p. 891-916, 2020.
- CAMPOS, Maria Malta *et al.* A qualidade da educação infantil: um estudo em seis capitais brasileiras. **Cadernos de pesquisa**, v. 41, n. 142, p. 20-54, 2011.
- COOK, David A.; BECKMAN, Thomas J. Current concepts in validity and reliability for psychometric instruments: theory and application. **The American journal of medicine**, v. 119, n. 2, p. 166. e7-166. e16, 2006.
- 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. Disponível em:  
<<https://journals.sagepub.com/doi/abs/10.1177/016146811011200303>>.
- CAMPBELL, F *et al.* Early childhood investments substantially boost adult health. **Science**, v. 343, n. 6178, p. 1478-1485, 2014. Disponível em:  
<<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4028126/>>.
- 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. Disponível em:  
<[https://www.nber.org/system/files/working\\_papers/w9495/w9495.pdf](https://www.nber.org/system/files/working_papers/w9495/w9495.pdf)>.
- 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.
- CICCHETTI, D. V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. **Psychological Assessment**, v. 6, n. 4, p. 284-290, 1994. Disponível em <  
[https://www.researchgate.net/publication/232556850\\_Guidelines\\_Criteria\\_and\\_Rules\\_of\\_Thumb\\_for\\_Evaluating\\_Normed\\_and\\_Standardized\\_Assessment\\_Instrument\\_in\\_Psychology](https://www.researchgate.net/publication/232556850_Guidelines_Criteria_and_Rules_of_Thumb_for_Evaluating_Normed_and_Standardized_Assessment_Instrument_in_Psychology)>.
- DEVELLIS, Robert F.; THORPE, Carolyn T. **Scale development: Theory and applications**. Sage publications, 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.
- DOHRENWEND, B. P. *et al.* Nonspecific psychological distress and other dimensions of psychopathology: Measures for use in the general population. **Arch Gen Psychiatry**,

v. 37, n. 11, p. 1229-1236, nov. 1980. Disponível em:  
<<https://pubmed.ncbi.nlm.nih.gov/7436685/>>.

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.

ECHEVARRÍA-GUANILO, Maria Elena; GONÇALVES, Natália; ROMANOSKI, Priscila Juceli. Propriedades psicométricas de instrumentos de medidas: bases conceituais e métodos de avaliação-Parte I. **Texto & Contexto-Enfermagem**, v. 26, 2018.

FEINSTEIN, Alvan R.; CICCHETTI, Domenic V. High agreement but low kappa: I. The problems of two paradoxes. **Journal of clinical epidemiology**, v. 43, n. 6, p. 543-549, 1990.

FLEISS, Joseph L.; LEVIN, Bruce; PAIK, Myunghee Cho. **Statistical methods for rates and proportions**. John Wiley & Sons, 2013.

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.

FREITAS, A. L. P.; RODRIGUES, S. G. A. Avaliação da confiabilidade de questionário: uma análise utilizando o coeficiente alfa de Cronbach In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 12, 2005, 07-09 nov, Bauru-SP. Anais... Bauru-SP: UNESP, 2005. Disponível em:  
<[https://simpep.feb.unesp.br/anais/anais\\_12/copiar.php?arquivo=Freitas\\_ALP\\_A%20avalia%E7%E3o%20da%20confiabilidade.pdf](https://simpep.feb.unesp.br/anais/anais_12/copiar.php?arquivo=Freitas_ALP_A%20avalia%E7%E3o%20da%20confiabilidade.pdf)>.

FUZZARI, Paulo Gabriel Guerche *et al.* Relatos de Grupos de Trabalho: pesquisando instrumentos de avaliação em cooperação com o setor público. **Revista Brasileira de Avaliação**, v. 11, n. 3 spe, p. 1-7, 2022.

GRIEP, Rosane Harter *et al.* Apoio social: confiabilidade teste-reteste de escala no Estudo Pró-Saúde. **Cadernos de Saúde Pública**, v. 19, p. 625-634, 2003.

GULLO, Dominic F. Improving instructional practices, policies, and student outcomes for early childhood language and literacy through data-driven decision making. **Early Childhood Education Journal**, v. 41, p. 413-421, 2013.

HAIR JR, Joseph F. *et al.* **Partial least squares structural equation modeling (PLS-SEM) using R: A workbook**. Springer Nature, 2021.

HARRIS, Monica J.; ROSENTHAL, Robert. Mediation of interpersonal expectancy effects: 31 meta-analyses. **Psychological bulletin**, v. 97, n. 3, p. 363, 1985.

HECKMAN, J. J. Skill formation and the economics of investing in disadvantaged children. **Science**, v. 312, n. 5782, p. 1900-1902, 2006. Disponível em:  
<<https://www.science.org/doi/10.1126/science.1128898>>.

- 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. Disponível em: <[https://www.anpec.org.br/encontro/2022/submissao/files\\_I/i12-00786895c8741d1da9233d24174f0c2d.pdf](https://www.anpec.org.br/encontro/2022/submissao/files_I/i12-00786895c8741d1da9233d24174f0c2d.pdf)>.
- JONES, S. M., & BOUFFARD, S. M. Social and emotional learning in schools: From programs to strategies. **Social Policy Report**, v. 26, n. 4, p. 1-33, 2012. Disponível em: <<https://files.eric.ed.gov/fulltext/ED540203.pdf>>.
- KOO, Terry K.; LI, Mae Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. **Journal of chiropractic medicine**, v. 15, n. 2, p. 155-163, 2016.
- LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, p. 159–174, 1977. Disponível em: <<https://www.jstor.org/stable/2529310>>.
- LIMA, Isabel Christinie Fontgalland Rodrigues de. **Ensaio sobre a educação infantil: Análises com os pais, as professoras e as crianças**. 2022. 125 f. Dissertação de Mestrado - Programa de Pós-Graduação em Economia, Universidade Federal do Ceará, Fortaleza, 2022.
- LOCASALE-CROUCH, Jennifer *et al.* Pre-kindergarten teachers' use of transition practices and children's adjustment to kindergarten. **Early childhood research quarterly**, v. 23, n. 1, p. 124-139, 2008.
- MACPHEE, D. Manual: Knowledge of infant development inventory. **Unpublished manuscript, University of North Carolina**, 1981.
- MAIA, Joviane Marcondelli Dias; WILLIAMS, Lucia Cavalcanti de Albuquerque. Fatores de risco e fatores de proteção ao desenvolvimento infantil: uma revisão da área. **Temas em psicologia**, v. 13, n. 2, p. 91-103, 2005.
- MALDONADO-CARREÑO, C. *et al.* Measuring the quality of early childhood education: Associations with children's development from a national study with the IMCEIC tool in Colombia. **Child development**, v. 93, n. 1, p. 254-268, 2022. Disponível em: <<https://srd.onlinelibrary.wiley.com/doi/full/10.1111/cdev.13665>>.
- MAROPE, Priscilla Toka Mmantsetsa; KAGA, Yoshie. **Investing against evidence: The global state of early childhood care and education**. UNESCO Publishing, 2015.
- MARX, Robert G. *et al.* A comparison of two time intervals for test-retest reliability of health status instruments. **Journal of clinical epidemiology**, v. 56, n. 8, p. 730-735, 2003.
- NUNNALLY, Jum C. *et al.* Bernstein. Ih (1994). **Psychometric theory**, v. 3, 1994.
- O'LEARY, Matt. **Classroom observation: A guide to the effective observation of teaching and learning**. Routledge, 2020.

ONES, D. S.; VISWESVARAN, C.; REISS, A. D. Role of social desirability in personality testing for personnel selection: The red herring. **Journal of Applied Psychology**, v. 81, n. 6, p. 660-679, 1996. Disponível em: <<https://experts.umn.edu/en/publications/role-of-social-desirability-in-personality-testing-for-personnel->>.

PAULHUS, D. L. Socially desirable responding: The evolution of a construct. In: BRAUN, H. I.; JACKSON, D. N.; WILEY, D. E. (Eds.). The role of constructs in psychological and educational measurement. **Lawrence Erlbaum Associates Publishers**, p. 49-69, 2002. Disponível em: <<https://psycnet.apa.org/record/2001-05107-004>>.

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>>

RAIKES, Abbie. Using Early Childhood Data and Measurement to Leverage Change: Are We Making Progress?. **Journal of Research in Childhood Education**, p. 1-11, 2023. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/02568543.2023.2248223>>.

RAIKES, A. *et al.* Examination of school readiness constructs in Tanzania: Psychometric evaluation of the MELQO scales. **Journal of Applied Developmental Psychology**, v. 62, p. 122-134, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0193397318301643>>.

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. Disponível em: <<https://journals.sagepub.com/doi/10.3102/01623737024004267>>.

REYNOLDS, A. *et al.* Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. **JAMA**, v. 285, n. 18, p. 2339-2346, mai., 2001. Disponível em: <<https://jamanetwork.com/journals/jama/fullarticle/193816>>.

SALMOND, Susan S. Evaluating the reliability and validity of measurement instruments. **Orthopaedic Nursing**, v. 27, n. 1, p. 28-30, 2008.

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. Disponível em: <[http://scielo.iec.gov.br/scielo.php?script=sci\\_arttext&pid=S1679-49742017000300649&lng=pt&nrm=iso](http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742017000300649&lng=pt&nrm=iso)>.

- SCHEINHART, L. J. *et al.* Lifetime effects: The HighScope Perry Preschool study through age 40. **HighScope Press**, 2005. Disponível em: <[https://nieer.org/wp-content/uploads/2014/09/specialsummary\\_rev2011\\_02\\_2.pdf](https://nieer.org/wp-content/uploads/2014/09/specialsummary_rev2011_02_2.pdf)>.
- 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).
- SHROUT, Patrick E.; FLEISS, Joseph L. Intraclass correlations: uses in assessing rater reliability. **Psychological Bulletin**, v. 86, n. 2, p. 420-428, 1979. Disponível em: <<https://psycnet.apa.org/fulltext/1979-25169-001.pdf>>
- SIM, Julius; WRIGHT, Chris C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. **Physical therapy**, v. 85, n. 3, p. 257-268, 2005.
- SWELLER, J. Cognitive load theory, learning difficulty, and instructional design. **Learning and Instruction**, v. 4, n. 4, p. 295, 1994. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/0959475294900035>>.
- SU, Y. *et al.* Preschool quality and child development in China. **Early childhood research quarterly**, v. 56, p. 15-26, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0885200621000107?via%3Dihub>>.
- THORPE, Karen *et al.* Meal provision in early childhood education and care programs: Association with geographic disadvantage, social disadvantage, cost, and market competition in an Australian population. **Social Science & Medicine**, v. 312, p. 115317, 2022.
- 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.
- VANDELL, Deborah Lowe *et al.* Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. **Child development**, v. 81, n. 3, p. 737-756, 2010.
- VOLPE, Robert J. *et al.* Observing students in classroom settings: A review of seven coding schemes. **School Psychology Review**, v. 34, n. 4, p. 454-474, 2005.
- UNESCO, UNICEF, Brookings Institution and the World Bank. Overview Measuring Early Learning Quality and Outcomes. 2017. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000248053>>
- WORLD BANK. Measuring Early Learning Quality and Outcomes (MELQO) Framework to Early Childhood Education Settings in Colombia. 2017.

YOSHIKAWA, Hirokazu *et al.* Investing in our future: The evidence base on preschool education. **Society for Research in Child Development**, 2013.

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.

ZASLOW, Martha *et al.* Quality measurement in early childhood settings. **Baltimore, MD: Brookes**, 2011.

## APÊNDICE A

**Quadro A1** – Itens da EAPI que compõem as dimensões, seções e facetas.

Dimensão	Seção	Faceta	Instrumento*	Item
Alimentação	Gestão	-	OD	Momentos de alimentação foram observados?
			OD	Cardápio atualizado com as refeições do dia;
			OD	Alimentação servida atende ao cardápio;
			OD	Guardanapos acessíveis para uso das crianças;
			OD	Mesas e cadeiras estão limpas quando as crianças chegam para a refeição;
	Infraestrutura	-	OD	Mesas e cadeiras adequadas ao tamanho das crianças;
			OD	Mesas possuem toalhas ou jogos americanos;
			OD	Organização do espaço permite a mobilidade das crianças;
			OD	Lixeira com tampa e pedal na altura das crianças;
	Práticas	-	OD	Práticas de lavagem das mãos;
			OD	Oportunidades de aprendizagens que envolvem práticas sociais de alimentação e nutrição;
			OD	Cardápio disponível à altura das crianças;
Currículo, interações e práticas pedagógicas	Organização dos tempos, espaços e materiais	Organização dos espaços	OD	Materiais diversos estão organizados no espaço da sala em cantos, áreas ou outra denominação de diferentes setores (além da leitura, por exemplo, o teatro, a fantasia, as coleções, os jogos e brinquedos, os blocos, as artes) - pelo menos três áreas organizadas e identificadas;
			OD	Produções feitas pelas crianças estão expostas na sala de referência da turma;
			OD	Produções feitas pelas crianças estão expostas em diversos espaços da UE;
			OD	Objetos pessoais e/ou fotos da criança e/ou da família estão presentes na sala de referência;
			OD	Textos, listas, cartazes ou informações escritas nas paredes estão na altura dos olhos das crianças;
			OD	Cartazes ou folhas impressas com ilustrações da rotina estão nas paredes na altura dos olhos das crianças;
		Organização dos materiais	OD	Utensílios para escrita e artes;
			OD	Riscadores fáceis de segurar;
			OD	Faz de conta;
			OD	Jogos;
			OD	Materiais não estruturados/de largo alcance/recicláveis;
			OD	Repertório artístico-cultural e científico de diferentes origens étnico-raciais;

	Organização dos tempos	OD	Livros de histórias;	
		OD	Materiais para pesquisa;	
		OD	Crianças vivenciam tempos de espera, de maneira contínua ou entre as atividades, sem nenhuma atividade específica ou sem possibilidade de escolha para uma nova atividade ou em atividade apenas para passar o tempo (ex.: assistir TV).	
		OD	Durante a rotina, o ritmo das crianças foi respeitado;	
		OD	Na organização do trabalho com o grupo de crianças, a(o) professora(or) prevê DUAS OU MAIS formas de organização que incluem:	
		OD	Tempo total de oportunidades de aprendizagens oferecidas em espaços externos da EU.	
		OD	Acompanhamento das crianças por um adulto.	
	Planejamento e currículo	Currículo	EP	O quanto você conhece os direitos de aprendizagem, campos de experiência e objetivos de aprendizagem e desenvolvimento da Base Nacional Comum Curricular?
			EP	Você considera que sua prática é realizada de acordo com o proposto nessa organização curricular?
		Planejamento	OD	Utiliza algum documento curricular oficial como base para elaboração do planejamento?
			OD	Você tem um registro de planejamento das atividades a serem desenvolvidas hoje com as crianças? (Se sim, peça acesso ao documento)
			OD	Com que frequência o planejamento é realizado?
			OD	As aprendizagens e o desenvolvimento das crianças são registrados continuamente? (Se sim, peça à(o) professora(or) que lhe mostre esse material)
			OD	Com que frequência o registro é realizado?
	OD	As crianças têm suas produções e registros de atividades organizados?		
Práticas de ampliação do repertório por meio do brincar	-	OD	Oportunidades de aprendizagens que envolvem práticas sociais com espaços, tempos, objetos e suas relações;	
		OD	Oportunidades de aprendizagens que envolvem práticas sociais com a leitura e a escrita;	
		OD	Oportunidades de aprendizagens que promovam a participação em práticas de oralidade;	
		OD	Leitura de livros de histórias pela(o) professora(or);	
		OD	Oportunidades de aprendizagens que ampliam a expressão e a criação por meio das linguagens plásticas;	
		OD	Oportunidades de aprendizagens que ampliam a expressão e a criação por meio de teatro, dança e/ou música;	
		OD	Oportunidades de aprendizagens que promovem os movimentos amplos e gestos das crianças em jogos e brincadeiras;	
		OD	Oportunidades de aprendizagens que ampliam a relação das crianças com as tecnologias digitais;	
OD	Momentos de brincadeira livre;			

			OD	Oportunidades de aprendizagens que envolvam as crianças em atividades de observação e investigação do mundo físico e da natureza;
			OD	Oportunidades de aprendizagens relacionadas à participação das crianças na organização das atividades cotidianas;
			OD	Oportunidades de aprendizagens que abordem aspectos relacionados à educação étnico-racial;
	Práticas para acolhimento e gestão de conflitos	-	OD	Estratégias verbais de gestão dos conflitos e dos comportamentos das crianças pela(o) professora(or);
			OD	Estratégias de acolhimento afetivo às crianças;
			OD	Professora(or) atendeu individualmente às crianças, a depender das necessidades e particularidades;
			OD	Presença de interações verbais negativas;
			OD	Presença de interações físicas negativas;
			OD	Presença de interações verbais negativas em relação a grupos étnico-raciais ou segmentos sociais;
			OD	Envolvimento das crianças;
	Práticas para o cuidado de si, o bem-estar e a saúde	-	OD	Oportunidades de aprendizagens que envolvem práticas sociais de alimentação e nutrição;
			OD	Oportunidades de aprendizagens relacionadas aos procedimentos de higiene (uso do banheiro);
			OD	Oportunidades de aprendizagens relacionadas aos momentos de sono e descanso das crianças;
			OD	Durante o momento de sono, alguma criança manifestou o desejo de ficar acordada, ficou acordada durante todo o período ou acordou antes das demais?
			OD	Durante a observação, alguma criança manifestou a necessidade de dormir ou repousar fora de um horário preestabelecido?
			OD	Crianças têm acesso livre e autônomo à água para beber;
			OD	Crianças podem beber água quando desejam, mesmo que precisem pedir para um adulto;
			OD	Crianças têm acesso livre e autônomo às instalações sanitárias;
			OD	Crianças podem ir ao banheiro quando desejam, mesmo que precisem pedir para um adulto;
Diversidade funcional	-	-	ED	Há professoras(es) com formação específica para a inclusão de crianças com deficiência? Quantas(os)?
			ED	Há crianças com deficiências, TEA ou altas habilidades?
			ED	Há quantos profissionais na unidade dedicados somente ao trabalho com crianças com deficiência?
			ED	Há uma sala de Atendimento Educacional Especializado (espaço adaptado para atividades com crianças com deficiências) em sua UE?
			EP	Você fala algum outro idioma além de português? Qual?
			OD	Piso tátil direcional do ponto de ônibus mais próximo até a entrada da UE;
			OD	Sinalização sonora;
			OD	Existe campainha ou interfone acessível a pessoas em cadeira de rodas e crianças menores;

			OD	Piso tátil direcional nas dependências da UE em pátios e espaços amplos;
			OD	Piso tátil de alerta na identificação de obstáculos e perigos potenciais;
			OD	Vias adequadas a crianças com deficiência ou mobilidade reduzida;
			OD	Portas com vão livre de no mínimo 80 cm;
			OD	Rampas como alternativa para degraus e escadas;
			OD	Corrimãos e grades de proteção para rampas e escadas;
			OD	Elevador ou plataformas elevatórias entre andares;
			OD	Piso, paredes e móveis possuem cores contrastantes;
			OD	Presença de sinais/pictogramas complementando informações escritas;
			OD	Dependências adequadas a crianças com deficiência ou mobilidade reduzida;
			OD	Mobiliário adequado para a aproximação e uso por crianças em cadeira de rodas;
			OD	Banheiro adaptado para crianças com deficiência ou mobilidade reduzida;
			OD	Adaptações nos materiais da sala de referência (livros em braille, materiais que permitam discriminação visual e/ou tátil, materiais que facilitam o manuseio de crianças com dificuldades de preensão, lápis com engrossador, etc).
			OD	Professora(or) atendeu individualmente às crianças, a depender das necessidades e particularidades;
Equipe e gestão	Equipe	Condições de trabalho	ED	Há espaço para reuniões, estudos, momentos de formação e planejamento?
			ED	Há banheiro de uso exclusivo dos profissionais com pia e vaso sanitário de tamanho adulto?
			EP	Você paga por alguma despesa relacionada às atividades ou à escola (brinquedos, riscantes etc.)?
			EP	Estou satisfeita(o) com a localização de meu trabalho;
			EP	A gestão da unidade conhece e implementa procedimentos que visam prevenir problemas de saúde das(os) professoras(es) e demais profissionais?
			OD	Banheiro adaptado para adultos com deficiência ou mobilidade reduzida;
		OD	O espaço é suficiente para as crianças participarem das atividades realizadas dentro da sala.	
		Formação	ED	Formações realizadas por assessores/institutos contratados pelo município;
			ED	Formações oferecidas por editoras;
			ED	Formações em contexto (horário coletivo);
			ED	Formações continuadas de programas estaduais;
			ED	Formações continuadas oferecidas pela secretaria municipal;
		EP	Você faz uso de alguma outra língua no trabalho com as crianças?	
Motivação e engajamento	EP	Estou satisfeita(o) com a posição que ocupo em minha carreira;		
	EP	Estou satisfeita(o) com o trabalho que realizo;		

	Percepção dos professores sobre formação	EP	É estimulante para mim a possibilidade de planejar as experiências e oportunidades que irei oferecer às crianças;	
		EP	Como tem sido sua participação na elaboração do documento político-pedagógico da unidade educacional? Caso não, por quê?	
		EP	Os momentos de cooperação entre professores, promovidos pela instituição, são fonte de inspiração para mim;	
		EP	Eu sinto que tenho a formação necessária para ser uma(um) boa(bom) professora(or) de Educação Infantil;	
		EP	Me sinto preparada(o) para educar crianças com deficiências, TEA ou altas habilidades;	
		EP	Eu sinto que sou formada(o) continuamente para aprimorar meu trabalho na Educação Infantil;	
	Gestão	Apoio à equipe	EP	Meu trabalho é desenvolvido colaborativamente com a gestão da EU;
			EP	A gestão busca proporcionar condições para o desenvolvimento do meu trabalho docente;
			EP	Eu sinto que há esforço da equipe da UE para acolher e formar os profissionais recém contratados;
			EP	Eu estou sobrecarregada(o) com o volume de trabalho que tenho a fazer;
			EP	Há momentos coletivos propostos e organizados pela gestão da unidade para reflexão, documentação e planejamento das práticas pedagógicas desenvolvidas com as crianças? Se sim, com que frequência?
			EP	Os momentos coletivos de organização e planejamento possuem objetivos claros para as discussões?
			EP	Você considera que a carga horária reservada para o planejamento individual é suficiente?
			Espaços coletivos	ED
		EP		Há momentos coletivos propostos e organizados pela gestão da unidade para reflexão, documentação e planejamento das práticas pedagógicas desenvolvidas com as crianças? Se sim, com que frequência?
		EP		Os momentos coletivos de organização e planejamento possuem objetivos claros para as discussões?
		Gestão de recursos materiais	ED	Há manutenção e reposição de equipamentos mensalmente?
			ED	Há um plano de manutenção e reforma das edificações para planejar e executar ações corretivas preventivas? (Peça à(ao) diretora(or) que lhe mostre um documento)
			ED	É realizada conservação predial anualmente?
		Projeto Político Pedagógico	ED	Você organiza espaços coletivos para a participação da equipe de profissionais na elaboração do projeto político-pedagógico da unidade educacional? (Se sim, perguntar quantos participam)

			EP	Como tem sido sua participação na elaboração do documento político-pedagógico da unidade educacional? Caso não, por quê?
			EP	Sua construção é realizada de maneira coletiva;
			EP	Sua construção é realizada de maneira consultiva;
			EP	É atualizado periodicamente;
			EP	Fica disponível para consulta;
			EP	É ponto de partida para o planejamento das atividades;
Infraestrutura	Equipamentos	Água	ED	A escola utiliza alguma estratégia de tratamento/filtragem de água para torná-la potável? Se sim, com que frequência?
			OD	Utensílios para beber água limpos, individualizados e não descartáveis;
		Sanitários	OD	São adequados ao tamanho das crianças;
			OD	Contam com assento de vaso sanitário;
			OD	São limpos;
			OD	Contam com dispensadores de sabonete;
			OD	Contam com dispensadores de papéis;
			OD	Contam com cesto de lixo;
			OD	Têm sabonete nos dispensadores;
			OD	Têm papel toalha nos dispensadores;
		Sono e conforto	OD	Têm papel higiênico nos dispensadores;
			OD	Os 3 itens acima são acessíveis às crianças;
	OD		Crianças têm acesso a uma área com materiais macios;	
	OD		Possui colchonetes suficientes e em bom estado;	
	Troca de fralda	OD	Roupa de cama limpa e de uso exclusivo da criança;	
		OD	Iluminação, ventilação e som adequados para o sono;	
	Estrutura predial	Ambiente externo	OD	Espaço suficiente para o sono;
			OD	Possuem bancada para troca de fraldas, acompanhada de colchonete (trocador) (para bebês e crianças menores);
			OD	A bancada para troca de fraldas é anexada à banheira ou cuba (para bebês e crianças menores);
			OD	Há área externa coberta que possibilite a brincadeira das crianças;
OD			Há área externa descoberta que possibilite a brincadeira das crianças;	
			OD	Há parque infantil (plástico ou madeira);
			OD	Há área com areia;
			OD	Bebedouro com água filtrada na altura das crianças;

	Ambiente interno	OD	Materiais lúdicos pontiagudos ou enferrujados;	
		OD	Brinquedos do pátio mal conservados;	
		OD	Cadeiras quebradas;	
		OD	Goteiras no telhado, infiltração ou buracos no teto;	
		OD	Portas ou janelas quebradas;	
		OD	O espaço é suficiente para as crianças participarem das atividades realizadas dentro da sala;	
		OD	Iluminação adequada;	
		OD	Ventilação adequada;	
		OD	A sala possui janelas na altura das crianças ou paredes baixas, de forma que as crianças consigam olhar para fora da sala;	
		OD	Crianças têm acesso a mobiliário adequado (mesa, cadeira ou outro suporte para desenho ou escrita – ex. flip chart) quando realizam ou querem realizar uma atividade como desenhar, pintar, escrever;	
		OD	Possui bancadas, prateleiras e/ou armários para a guarda de brinquedos, materiais, livros, e outros, da altura das crianças;	
		OD	Espelho amplo na altura das crianças;	
		Geral	ED	Há uma sala de leitura/biblioteca?
			ED	Há espaço para reuniões, estudos, momentos de formação e planejamento?
	ED		Há banheiro de uso exclusivo dos profissionais com pia e vaso sanitário de tamanho adulto?	
	ED		Há um depósito para materiais de uso pedagógico em sua UE?	
	OD		Espaço em mau estado de conservação (pintura das paredes, móveis, metais etc.);	
	OD		Objetos empacotados / mal-armazenados na escola;	
	Materiais	-	OD	Objetos da sala de referência em mau estado de conservação;
			OD	Utensílios para escrita e artes;
OD			Riscadores fáceis de segurar;	
OD			Faz de conta;	
OD			Jogos;	
OD			Materiais não estruturados/de largo alcance/recicláveis;	
OD			Repertório artístico-cultural e científico de diferentes origens étnico-raciais;	
OD			Livros de histórias;	
OD			Materiais para pesquisa;	
OD			Número de livros de literatura;	
Segurança	-	-	OD	Número de materiais para pesquisa;
			EP	Você recebeu alguma formação sobre primeiros socorros nos últimos 12 meses?

		OD	Pisos quebrados ou desnivelados;
		OD	Materiais lúdicos pontiagudos ou enferrujados;
		OD	Áreas externas com pragas urbanas;
		OD	Fiação exposta ou mal conservada;
		OD	Tomadas elétricas sem proteção;
		OD	Poços/buracos descobertos;
		OD	Área de circulação de veículos sem isolamento por grades ou muros;
		OD	Extintores de incêndio ausentes;
		OD	Extintores de incêndio vencidos;
		OD	Pátio com muito lixo ou pedras;
		OD	Outras condições que possam vir a ferir as crianças;
		OD	Calhas e “canaletas” para escoamento da chuva obstruídas;
		OD	Corrimãos e grades de proteção para rampas e escadas;
		OD	Acompanhamento das crianças por um adulto;

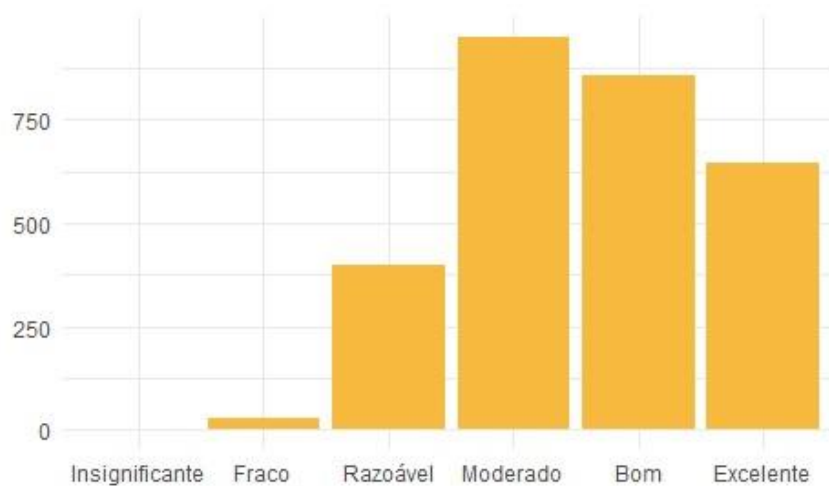
\*ED: Entrevista com diretor(a);

EP: Entrevista com professor(a);

OD: Observação direta.

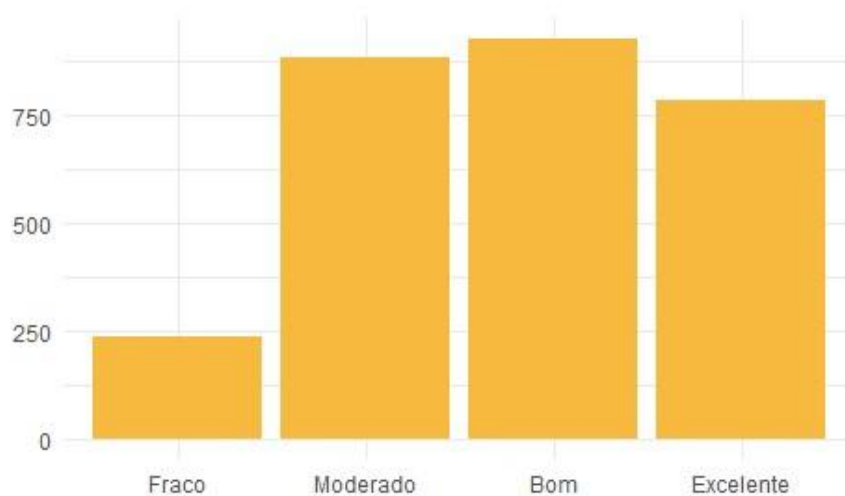
## APÊNDICE B

**Gráfico 2** – Distribuição dos coeficientes Kappa das turmas por dimensão.



Fonte: Elaborado a partir dos dados da EAPI Contínua.

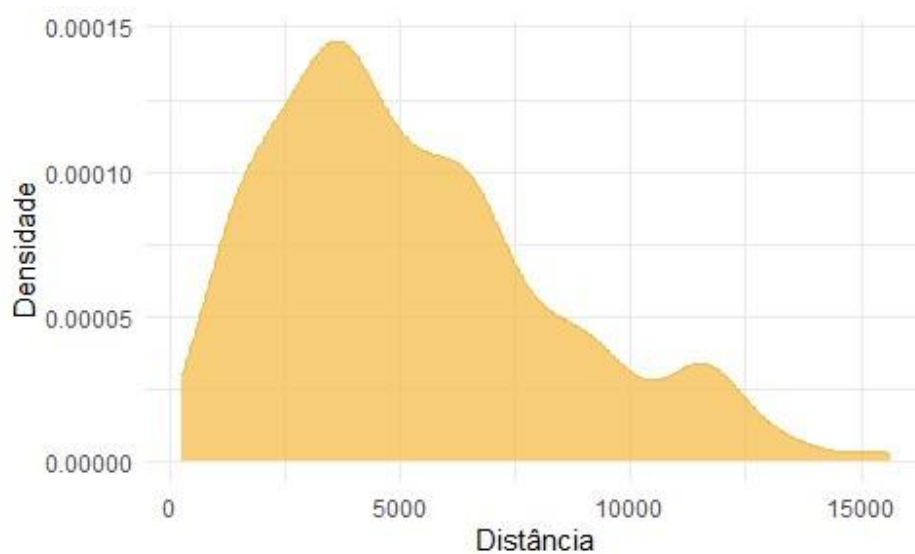
**Gráfico 3** – Distribuição dos coeficientes ICC das turmas por dimensão.



Fonte: Elaborado a partir dos dados da EAPI Contínua.

**Gráfico 4** – Distribuição dos coeficientes alfa de Cronbach das turmas por dimensão.

Fonte: Elaborado a partir dos dados da EAPI Contínua.

**Gráfico 5** – Densidade das distâncias percorridas pelos observadores de casa para a Unidade Educacional.

Fonte: Elaborado a partir dos dados da EAPI Contínua.