

DOI: 10.46667/renbio.v16i1.904

## CORRELAÇÃO E CAUSALIDADE EM AULAS DE BIOLOGIA BASEADAS NO ENSINO POR INVESTIGAÇÃO

### CORRELATION AND CAUSATION IN INQUIRY-BASED TEACHING BIOLOGY CLASSES

### CORRELACIÓN Y CAUSALIDAD EN LAS CLASES DE BIOLOGÍA BASADAS EN LA INDAGACIÓN

*Beatriz Busin Campos<sup>1</sup>, Máira Batistoni e Silva<sup>2</sup>*

#### Resumo

Correlação e causalidade são conceitos importantes para o entendimento de conhecimentos científicos, porém são comumente confundidos. Desenvolvemos e aplicamos uma sequência didática (SD) para investigar como estudantes do Ensino Médio relacionam variáveis na análise de dados e avaliar como uma SD que discute correlação e causalidade ajuda na construção de explicações baseadas em evidências. A SD investigativa abordou a microbiota intestinal e, durante sua aplicação, verificamos que os estudantes costumam inferir causalidade através de dados que demonstram correlações, muitas vezes utilizando conhecimentos prévios para justificar tais inferências. Também ponderamos que, por a SD ter sido aplicada remotamente, a interação entre pares, fundamental na abordagem investigativa, não foi efetiva, explicando a dificuldade de os estudantes reconhecerem e superarem os erros cometidos.

**Palavras-chave:** Ensino por investigação; Conhecimentos prévios; Microbiota intestinal; Ensino Médio.

#### Abstract

Correlation and causation are important concepts for the understanding of scientific knowledge, however they are commonly mistaken. We developed and applied a teaching sequence (TS) to investigate how high school students relate different variables in data analysis and to evaluate how a TS that discusses correlation and causation helps in the construction of evidence-based explanations. The investigative TS developed addressed the intestinal microbiota and, during its application, we found out that students usually infer causation through data that demonstrate correlations, often using prior knowledge to justify such inferences. We also consider that as the TS was applied remotely the interaction between peers, fundamental in the investigative approach, was not effective, explaining students' difficulty in recognizing and overcoming their mistakes.

**Keywords:** Inquiry-based teaching; Preview knowledges; Intestinal microbiota; High School.

---

<sup>1</sup> Mestranda no Programa de Pós-Graduação Interunidades em Ensino de Ciências - Universidade de São Paulo (USP) São Paulo, SP - Brasil. Professora de ciências - Maple Bear Alto de Pinheiros. São Paulo, SP - Brasil. **E-mail:** [bibicampos@gmail.com](mailto:bibicampos@gmail.com)

<sup>2</sup> Doutora em Educação - Universidade de São Paulo (USP). São Paulo, SP - Brasil. Professor Doutor - Universidade de São Paulo (USP). São Paulo, SP - Brasil. **E-mail:** [mbatistoni@usp.br](mailto:mbatistoni@usp.br)

### Resumen

Correlación y causalidad son conceptos importantes para la comprensión del conocimiento científico, pero comúnmente se confunden. Desarrollamos y aplicamos una secuencia didáctica (SD) para investigar cómo los estudiantes de secundaria relacionan variables en el análisis de datos y evaluar cómo una SD que conversa correlación y causalidad ayuda en la construcción de explicaciones basadas en evidencia. La SD desarrollada abordó la microbiota intestinal y, durante su aplicación, comprobamos que los estudiantes suelen inferir causalidad a través de datos que demuestran correlaciones, muchas veces utilizando conocimientos previos para justificar tales inferencias. También consideramos que dado que se aplicó a distancia, la interacción entre pares no fue efectiva, lo que explica la dificultad de los estudiantes para reconocer y superar sus errores.

**Palabras clave:** Educación basada en la indagación; Conocimiento previo; Microbiota intestinal; Escuela secundaria.

## 1 Introdução

A compreensão adequada das ciências e suas tecnologias depende do conhecimento que cada pessoa construiu, em grande parte, nas situações de educação formal, ou seja, na instituição escolar, nos diferentes níveis de ensino. Assim, a ciência que é ensinada na escola precisa contribuir para que estudantes possam, em sua vida cotidiana, articular conhecimentos para a tomada de decisões.

Considerando a ciência como uma atividade construída socialmente a partir de práticas específicas que são sustentadas pelo compartilhamento de normas e linguagem que lhe são próprias (LONGINO, 2018; KUHN, 2020), nos últimos anos vem crescendo o número de pesquisas que abordam aspectos epistemológicos no ensino de ciências, destacando-se a ideia de que este não deve se preocupar somente com a aquisição de conceitos, mas também deve possibilitar que os alunos conheçam e se apropriem das práticas sociais da comunidade científica (KELLY, 2008; OSBORNE, 2016; KELLY & LICONA, 2018).

Não é incomum nos depararmos com a afirmação de que a explicação é o objetivo mais importante da ciência (STREVEN, 2013). Considerando seu importante papel na ciência, pesquisadores defendem que, no ensino, estratégias didáticas fomentadoras da construção e da compreensão de explicações sejam cruciais para o entendimento dos fenômenos e da natureza da ciência (CHINN & MALHOTRA, 2002; DUSCHL & GRANDY, 2008).

A explicação pode ser definida como uma ação cuja intenção é esclarecer algo ou torná-lo mais inteligível (NORRIS et al., 2005), ou ainda como um conjunto de proposições que deixa os fenômenos mais compreensíveis, estabelecendo informações sobre relações de dependência entre o fenômeno a ser explicado e as proposições que compõem a explicação (McCAIN, 2015). Mesmo que genéricas, essas definições são amplas o suficiente para contemplar os diversos modelos de explicações presentes na prática dos cientistas. Braaten e Windschitl (2011), por exemplo, analisaram diversos trabalhos de filosofia da ciência que buscavam examinar a estrutura e o papel das explicações na ciência e, assim, compilaram cinco modelos de explicação. Dentre os cinco modelos mapeados pelos autores, destacamos dois mais relevantes para o contexto do ensino das ciências e desta pesquisa: o estatístico-probabilístico (explicação

derivada de tendência ou padrão nos dados disponíveis sem explicitar a causa do evento - um exemplo é a correlação) e o causal (explicação derivada de tendência ou padrão nos dados explicitando a causa do evento - a causalidade).

A correlação estatística é uma medida da força e da direcionalidade de uma relação entre variáveis (DEVROOP, 2000). Entende-se que duas variáveis estão correlacionadas quando a variação de uma é acompanhada, em média, por variação na outra, sendo que essas variações podem se dar com intensidades e direções diferentes (CURRAN-EVERETT, 2010). Dessa maneira, ao evidenciar uma correlação entre variações em x e y, entende-se que são possíveis os seguintes cenários: i. alterações em x resultam em alterações em y (x causa y); ii. alterações em y resultam em alterações em x (y causa x); iii. alterações em z resultam em alterações em x e y (z causa x e z causa y); e iv. a correlação é fruto do acaso (z causa x e w causa y) (CURRAN-EVERETT, 2010; BARROWMAN, 2014).

Já a causalidade (também conhecida como relação de causa e efeito) é a agência que conecta uma variável (a causa) à outra (o efeito), em que a primeira é responsável pela segunda e a segunda depende da primeira. Isso posto, é importante afirmar que a existência de correlação não permite concluir sobre como as variáveis se relacionam e, conseqüentemente, não são suficientes para a elaboração de previsões precisas, mas ela pode justificar a importância de estudos sobre a natureza da relação entre as variáveis.

Muito embora a distinção entre a natureza desses modelos de explicação seja prática corriqueira no contexto científico, frequentemente as correlações são divulgadas de maneira equivocada por diversos meios de comunicação em massa (mídia corporativa, alternativa, redes sociais etc.) e a correlação demonstrada acaba sendo compreendida como uma causalidade pela população (EDITORIAL, 2001). Na educação formal, também é comum a dificuldade dos estudantes de distinguir esses dois modelos de explicação (MICHAEL, 2007; SLOMINSKI et al., 2019), embora sejam considerados conceitos importantes das dimensões de credibilidade da ciência e, portanto, para a compreensão sobre a natureza da ciência (ALLCHIN, 2013)

Interpretar errado uma relação causal ou identificá-la onde não existe pode resultar na tomada de ações não efetivas, na perpetuação de práticas prejudiciais e em perder a oportunidade de implementar alternativas benéficas (BARROWMAN, 2014). Isso pode ocorrer em muitos campos da vida humana, como em avaliações de ações governamentais, do impacto de novas tecnologias ou da eficácia de novos tratamentos médicos.

Tomemos como exemplo a polêmica acerca dos alimentos denominados ultraprocessados<sup>3</sup>: de um lado, alguns grupos de especialistas em saúde pública atribuem a esses alimentos um papel importante no aumento do risco de desenvolver problemas de saúde; por outro lado, pesquisadores da área de ciência e tecnologia de alimentos e lobistas da indústria alimentícia consideram a classificação imprecisa e afirmam que esse tipo de alimento garante a ingestão diária de energia e nutrientes necessários à grande parte da população mundial e que o consumo desse tipo de alimento é apenas um entre vários fatores a serem ponderados na explicação dos problemas de saúde da população (ZORZETTO, 2018).

Em 2016, uma pesquisa se baseou na avaliação de informações sobre 8451 pessoas divididas em grupos que consumiam diferentes quantidades de alimentos ultraprocessados. Após nove anos de acompanhamento, notou-se que as pessoas que consumiam mais alimentos ultraprocessados apresentaram 26% de chance de estarem acima do peso considerado saudável. Esse resultado expressa uma correlação positiva entre consumo de alimentos ultraprocessados e sobrepeso e, corretamente, os pesquisadores evitavam afirmar que os alimentos ultraprocessados provocam sobrepeso, dado que ainda se desconhecia o processo que relaciona essas duas variáveis. Os dados de pesquisas desse tipo não nos mostram se o que causa o problema de saúde é o grau de processamento ou a elevada ingestão de açúcares e gorduras associadas a esses alimentos. A correlação observada, portanto, não trazia evidências suficientes para desbancar o lobby de grandes multinacionais alimentícias que questionam a classificação dos alimentos de acordo com seu grau de processamento.

Já em 2019, foram publicados os resultados de um outro tipo de pesquisa, baseada em um ensaio clínico controlado, na qual um grupo de 10 pessoas recebeu uma dieta com 80% dos alimentos ultraprocessados, enquanto outro grupo semelhante recebeu uma dieta com 80% de alimentos *in natura* ou minimamente processados. A quantidade de energia e nutrientes era idêntica nas duas dietas. Após 15 dias de observação, as pessoas do primeiro grupo engordaram em média 1 kg, ao passo que as pessoas que tiveram uma dieta rica em alimentos *in natura*, emagreceram em média 1 kg (HALL et al., 2019). Esses resultados sustentam uma relação causal entre o processamento dos alimentos e o ganho de peso, apontando um alvo preciso para políticas públicas com vistas à redução do consumo de alimentos ultraprocessados e promoção da saúde da população.

Considerando a relevância das correlações e das relações causais para a construção do conhecimento científico e para decisões em questões sociocientíficas, entendemos que estratégias didáticas fomentadoras da construção e da interpretação dessas relações entre variáveis são cruciais, não apenas para o adequado entendimento dos fenômenos em si, mas

---

<sup>3</sup> Alimentos que passaram por maior processamento industrial e que normalmente possuem alta adição de açúcares, gorduras e substâncias sintetizadas em laboratório que aumentam sua durabilidade ou lhe conferem cor e sabor.

DOI: 10.46667/renbio.v16i1.904

também para a apropriação dos tipos de raciocínios envolvidos na construção de explicações científicas.

Nossa defesa para o ensino de ciências e biologia na educação básica não está restrita a expor os estudantes às características epistêmicas de cada modelo explicativo e esperar que os mesmos sejam capazes de declará-las em situações escolares. Ao invés disso, defendemos que os estudantes tenham oportunidades de analisar diferentes tipos de dados e construir explicações pautadas em evidências para que, engajados nessas práticas científicas, possam construir entendimentos sobre os diferentes tipos de relações entre as variáveis e as diferentes explicações delas derivadas.

Diante do exposto, interessa-nos: i. investigar como estudantes do Ensino Médio relacionam diferentes variáveis na análise de dados de um determinado fenômeno, e ii. avaliar como uma sequência didática que aborda explicitamente as relações de correlação e causalidade pode ajudar os estudantes a reconhecerem e a construir explicações baseadas em evidências.

## 2 Procedimentos Metodológicos

### 2.1 Planejamento da Sequência Didática

Para o desenvolvimento da pesquisa, elaboramos uma sequência didática (SD) baseada na abordagem didática do Ensino por Investigação (EnCI). De acordo com a revisão bibliográfica realizada por Minner et al. (2010), o EnCI pode ser sistematizado com base nos objetivos investigativos do *National Research Council* - NRC (2000). De acordo com esse documento, no EnCI os estudantes: i. são envolvidos em questões cientificamente orientadas; ii. dão prioridade às evidências para abordar questões cientificamente orientadas e propor explicações; iii. avaliam suas explicações à luz de proposições alternativas; e iv. comunicam e justificam as explicações propostas.

Para planejar a SD, utilizamos a estrutura do ciclo investigativo, um modelo instrucional fruto de revisão bibliográfica realizada por Pedaste et al. (2015). De acordo com os autores, o ciclo investigativo é composto pelas seguintes fases: Orientação - estimula a curiosidade sobre um tópico; Conceitualização - formula questões e/ou hipóteses baseadas no conhecimento prévio; Investigação - planeja e executa, por meio da exploração ou experimentação, a coleta e a análise de dados; Conclusão - propõe uma explicação com base nos dados analisados para responder a questão de investigação, retomando a hipótese ou previsão proposta na fase de conceitualização; Discussão - promove a comunicação e reflexão ao longo de todas as demais fases (PEDASTE et al., 2015).

A SD utilizada nesta pesquisa é composta por dois ciclos investigativos conectados, cuja temática principal é a microbiota intestinal de seres humanos – a comunidade de microrganismos (bactérias, fungos e protozoários) que habita o trato gastrointestinal humano. Essa escolha se justifica pelo fato de a microbiota intestinal estar relacionada a diversas variáveis, sem sabermos ao certo os processos que atuam nessas relações, tornando-a uma temática favorável para trabalharmos os modelos de correlação e causalidade.

Aplicamos a SD nas aulas de biologia da 3ª série do Ensino Médio de uma escola estadual na capital paulista. Tal aplicação ocorreu em novembro de 2020, durante a pandemia de COVID-19. Desta forma, a SD foi adaptada para o contexto do ensino remoto através da plataforma Google Meet, com um encontro semanal de uma hora de duração. Foram necessários três encontros síncronos para a aplicação completa da SD, além de quatro atividades realizadas de maneira assíncrona pelos estudantes. A seguir, apresentamos uma descrição das atividades que compuseram cada ciclo investigativo.

*1ª ciclo investigativo - buscando relações entre as características dos indivíduos e a microbiota*

Na atividade 1 da SD, os estudantes foram instruídos a assistir a uma videoaula<sup>4</sup> de 3min46s. Tratou-se de uma estratégia do tipo “sala de aula invertida”<sup>5</sup>, cujo propósito foi introduzir os estudantes à temática que seria estudada e apresentar a pergunta de investigação, contemplando a fase de orientação do ciclo investigativo. O vídeo apresentou definições de microrganismos, microbiota e sua importância fisiológica, assim como uma narrativa sobre como são realizadas pesquisas populacionais sobre microbiota. Por fim, foi apresentada a pergunta de investigação: “Existe relação entre as características dos indivíduos (sexo, % de gordura corporal, consumo diário de frutas, uso de antibiótico no último mês, hábito intestinal e consumo de café) e a composição da microbiota intestinal?”

A atividade 2 da SD contemplou a fase de investigação do ciclo investigativo e baseou-se na exploração de dados qualitativos e quantitativos extraídos do artigo *Population-level analysis of gut microbiome variation*<sup>6</sup> (FALONY et al., 2016). Individualmente, os alunos e as alunas receberam uma tabela contendo dados sobre características de 20 indivíduos e a composição da microbiota de cada um deles. Após analisarem essa tabela, os/as estudantes

---

<sup>4</sup> Link para acesso a videoaula: <https://youtu.be/a0x9LqJeQDE>.

<sup>5</sup> Por meio da metodologia da “sala de aula invertida” (do inglês, *flipped classroom*), antes da aula o/a estudante internaliza conceitos fundamentais por meio de atividades passadas pelo/a professor/a e posteriormente discute com os/as colegas os conhecimentos obtidos e tira possíveis dúvidas de conteúdo com a ajuda e orientação do/a professor/a.

<sup>6</sup> Análise em nível populacional da variação do microbioma intestinal, em tradução livre nossa para o português

deviam decidir se havia relação entre as diferentes características dos indivíduos e a composição da microbiota intestinal.

Na figura 1, vemos um trecho da tabela de dados que os/as estudantes analisaram. Cada coluna representa um indivíduo. As primeiras seis linhas apresentam fatores que poderiam estar correlacionados à composição da microbiota e as linhas abaixo representam gêneros de bactérias da microbiota intestinal. A letra P indica a presença dos gêneros na microbiota intestinal de cada um dos indivíduos. Os dados usados na tabela foram adaptados de Falony et al. (2016).

**Figura 1:** Trecho da tabela analisada na atividade 1.

Indivíduo	1	2	3	4	5	6	7	8
Sexo	Feminino	Masculino	Feminino	Feminino	Feminino	Masculino	Masculino	Feminino
% Gordura Corporal	Abaixo	Acima	Normal	Abaixo	Normal	Acima	Levemente Acima	Levemente Acima
Consumo de Frutas (semanal)	14	14	10	7	4	1	5	10
Uso de Antibióticos (no mês)	Não	Sim	Não	Não	Sim	Sim	Não	Não
Hábito intestinal (dias sem evacuar)	0	0	0	1	2	1	3	0
Consumo de Café	Frequente	Frequente	Frequente	Raramente	Raramente	Frequente	Frequente	Frequente
Bacteroides	P	P	P	P	P	P	P	P
Barnesiella	P		P	P		P		P
Bifidobacteriaceae gen1								
Bifidobacterium	P					P		P
Blautia	P			P			P	
Bryantella								
(...)								

Fonte: autoria própria

A atividade 3 da SD e última desse ciclo investigativo ocorreu no Google Meet e consistiu na apresentação e discussão da análise realizada pelos/as estudantes na atividade anterior, contemplando as fases de conclusão e discussão. Na discussão, buscou-se evidenciar que há variação da microbiota na população e demonstrar que porcentagem de gordura corporal, consumo de frutas, uso de antibióticos e hábito intestinal são fatores correlacionados à variação da microbiota, respondendo, portanto, a pergunta do primeiro ciclo investigativo.

### *2o ciclo investigativo - diferenciando correlação de causalidade*

Concebemos o segundo ciclo investigativo conectado ao primeiro, de tal forma que as atividades anteriores também constituem a fase de orientação deste novo ciclo.

Na atividade 4 da SD, os/as estudantes foram instruídos a assistir a uma segunda videoaula<sup>7</sup> de 3min58s. Novamente, tratou-se de uma estratégia do tipo sala de aula invertida, cujo propósito foi conceituar e diferenciar correlação e causalidade por meio da apresentação de diferentes gráficos com padrões semelhantes (correlações positivas) e da explanação sobre

<sup>7</sup> Link para acesso a videoaula: [https://youtu.be/\\_kbtZllc88](https://youtu.be/_kbtZllc88).

DOI: 10.46667/renbio.v16i1.904

como as variáveis estão relacionadas, evidenciando que a causalidade não pode ser determinada apenas pela análise da variação das variáveis.

A atividade 5 da SD contemplou a fase de investigação do ciclo investigativo e baseou-se na exploração dos dados anteriores para responder outra questão de investigação: “Os antibióticos matam bactérias e, por isso, há uma diminuição da diversidade da microbiota em indivíduos que utilizam esse tipo de medicação? Ou indivíduos que têm uma microbiota muito diversa têm mais chances de ter no organismo certas bactérias que protegem de doenças e, por causa dessas bactérias, necessitam de menor uso de antibióticos?”

Individualmente, os alunos e as alunas receberam duas afirmações: 1. Os antibióticos matam bactérias e, por isso, há uma diminuição da diversidade da microbiota em indivíduos que utilizam esse tipo de medicação; e 2. Indivíduos que têm uma microbiota muito diversa têm mais chances de ter no organismo certas bactérias que os protegem de doenças e, por causa dessas bactérias, necessitam de menor uso de antibióticos. Cada aluno deveria eleger uma das afirmações como correta e justificar sua escolha. Para ajudá-los nesse processo, foi-lhes solicitado que classificassem cada frase considerando as seguintes alternativas: a) Com certeza está correta; b) Desconfio que pode estar certa, mas os dados não permitem ter certeza; c) Desconfio que pode estar errada, mas os dados não permitem ter certeza; d) Com certeza está errada. Considerando os dados disponíveis para análise dos/das estudantes, as alternativas b) e c) são adequadas para as duas afirmações, visto que apresentam apenas uma correlação entre a composição da microbiota e o consumo de antibióticos.

A atividade 6 da SD ocorreu por meio do Google Meet com todo o grupo e consistiu na apresentação e discussão da análise realizada pelos estudantes na atividade 5, contemplando as fases de conclusão e discussão do segundo ciclo investigativo. Nesta atividade, os/as estudantes utilizaram o microfone e o chat para apresentar suas conclusões e justificativas.

A atividade 7 foi a última da SD e extrapola as fases dos ciclos investigativos, cujo propósito foi avaliar se as estratégias adotadas nos dois ciclos investigativos possibilitaram aos estudantes identificar corretamente as correlações e causalidades em diferentes tipos de dados e analisar explicações.

Na atividade 7, os/as estudantes reunidos em seis grupos em salas do Google Meet previamente criadas pela professora, deveriam ler dois textos de divulgação científica e determinar se eles estabeleciam correlação ou causalidade e se o faziam corretamente. O primeiro texto disponibilizado aos participantes da pesquisa possui a manchete “Microrganismos intestinais podem ajudar a diagnosticar câncer colorretal”<sup>8</sup> e traz a informação de que, por meio da composição da microbiota dos pacientes, é possível prever a presença ou não do câncer colorretal, apontando, de forma correta, uma correlação entre essas

<sup>8</sup> Fonte: Jornal da USP. Disponível em: <https://jornal.usp.br/ciencias/ciencias-da-saude/microrganismos-nas-fezes-podem-ajudar-a-diagnosticar-cancer-colorretal/>. Acesso em: 28 ago. 2022.

DOI: 10.46667/renbio.v16i1.904

duas variáveis. Já o segundo texto possui a manchete: “Cinco fatores que podem estar te fazendo engordar sem que você saiba o porquê”<sup>9</sup> e traz o exemplo de irmãs gêmeas que são muito parecidas, porém uma possui 41 quilos a mais do que a outra. Segundo o texto, a diferença de peso se deve à diferença na composição da microbiota intestinal das duas, uma vez que, quanto maior a diversidade da sua microbiota, mais magra é a pessoa. Este texto estabelece uma causalidade entre as variáveis, porém o faz de forma equivocada, dado que não apresenta nenhuma evidência de relação causal, trazendo apenas dados que demonstram uma correlação.

Depois de ler os textos, os/as estudantes deveriam escolher e justificar uma das seguintes alternativas: a) O primeiro e o segundo texto estabelecem causalidades de forma correta; b) O primeiro e o segundo texto estabelecem correlações de forma correta; c) O primeiro texto estabelece uma correlação de forma correta e o segundo estabelece uma causalidade equivocadamente; d) O primeiro texto estabelece uma causalidade de forma correta e o segundo estabelece uma correlação equivocadamente; e) O primeiro texto estabelece uma correlação de forma correta e o segundo texto estabelece uma correlação equivocadamente.

## 2.2 Coleta e análise de dados

A coleta de dados foi feita através do recolhimento dos registros individuais das atividades 2 e 5 (fase de investigação de cada ciclo, realizada individualmente e de forma assíncrona), e da gravação e transcrição do áudio e do chat das interações nos pequenos grupos durante a atividade 7.

Para a análise dos dados, utilizamos a Análise Textual Discursiva de Moraes e Galiazzi (2006), que consiste em técnicas para análise das comunicações a fim de identificar os núcleos de sentidos presentes em uma mensagem. Consideramos, como unidade de contexto, as atividades didáticas supracitadas (2, 5 e 7) e, como unidades de sentido, as frases (menor unidade textual com sentido) elaboradas como respostas pelos estudantes (atividade 2 e 5) ou os turnos de fala durante interação (atividade 7).

Após a definição do corpus, partimos para a etapa de categorização, para a qual optamos pela categorização a priori, com categorias baseadas nos pressupostos teóricos, conforme detalhado nos próximos parágrafos.

Para analisarmos as unidades de sentido, identificamos se o/a estudante mencionou uma das características abordadas nas atividades (sexo, porcentagem de gordura corporal, consumo de frutas, uso de antibiótico, hábito intestinal, consumo de café, câncer e sobrepeso) e se, ao mencioná-las, estabeleceu corretamente a existência ou não de relação entre a variável e a microbiota. Como critério, consideramos como causalidade se o/a estudante relacionou diretamente determinada característica com a microbiota, dando um sentido de que a

<sup>9</sup> Fonte: BBC News Brasil. Disponível em: <https://www.bbc.com/portuguese/geral-43955201>. Acesso em: 28 ago. 2022.

DOI: 10.46667/renbio.v16i1.904

característica individual determina a composição da microbiota. Da mesma forma, classificamos como correlação se o/a estudante identificou que a alteração de determinada característica é acompanhada com a alteração da microbiota, explicitando discursivamente que se trata de uma relação estatístico-probabilística fruto da análise populacional, ou se explicitou a possibilidade de diferentes cenários causais (x causa y; y causa x; z causa x e z causa y; z causa x e w causa y).

Para exemplificar, no Quadro 1 tomaremos o trecho da resposta de uma estudante para a atividade 2: “*Observei que entrevistados com % de gordura corporal acima da média possuíam potencialmente menos gêneros de bactérias. Aliado a isso, o fato dos dois indivíduos que ficam 3 dias sem evacuar terem menos gêneros também. Consumo de 7 a 14 frutas por semana evidentemente ajuda a microbiota intestinal a se desenvolver. O número mais disparado é o de uso de antibióticos, em que quase todos que não usaram no mês possuem consideravelmente mais gêneros presentes do que quem os usou, visto que obviamente os antibióticos matam as bactérias, sejam elas boas ou más.*”

**Quadro 1:** Exemplo de como se deu a categorização das unidades de sentido da atividade 2.

Unidade de sentido	Correlação	Causalidade
<i>Observei que entrevistados com % de gordura corporal acima da média possuíam potencialmente menos gêneros de bactérias.</i>	Aumento de gordura corporal é acompanhada potencialmente de redução dos gêneros de bactérias.	
<i>Aliado a isso, o fato dos dois indivíduos que ficam 3 dias sem evacuar terem menos gêneros também</i>	Aumento do tempo sem evacuação é acompanhado de redução dos gêneros de bactérias.	
<i>Consumo de 7 a 14 frutas por semana evidentemente ajuda a microbiota intestinal a se desenvolver.</i>		Consumo de frutas promove o desenvolvimento da microbiota.
<i>O número mais disparado é o de uso de antibióticos, em que quase todos que não usaram no mês possuem consideravelmente mais gêneros presentes do que quem os usou, visto que obviamente os antibióticos matam as bactérias, sejam elas boas ou más.</i>		Antibióticos matam bactérias, por isso, quem não usa possui mais gêneros de bactérias na microbiota.

Fonte: autoria própria

### 3 Resultados

A análise da atividade 2, na qual os/as estudantes deviam decidir se havia relação entre as diferentes características dos indivíduos e a composição da microbiota intestinal, gerou o Quadro 2, que mostra que 53,3% das relações estabelecidas pelos/as estudantes foram de causalidade e 46,6% foram de correlação, sendo que os dados oferecidos demonstravam apenas correlações. Tal resultado corrobora o que vem sendo apontado na literatura, pois pequenas demonstrações da coocorrência entre duas variáveis são frequentemente tomadas como evidências de uma relação causal (KUHN et al., 2015), o que torna bastante comum a inferência de causalidades através de dados que demonstram correlações (DEVROOP, 2000).

A distribuição dos dois tipos de respostas também revela a potencialidade da atividade proposta, visto que os estudantes tiveram que lidar com dados genuínos e suficientemente complexos para possibilitar a proposição de explicações. A constituição de um ambiente de aprendizagem no qual os estudantes possam pensar, errar, refletir sobre o erro e depois tentar um acerto é importante para a construção de novos conhecimentos e deve ser valorizado na abordagem do Ensino por Investigação (CARVALHO, 2013).

Os/as estudantes que estabeleceram as correlações corretamente se restringiram a descrever o que foi observado na tabela, como nas respostas seguintes: “A maioria dos indivíduos que ficaram de 2 a 3 dias sem evacuar possuem menor diversidade de microbiotas.”, “Indivíduos que evacuaram todos os dias possuem maior diversidade, entre 9 a 16 microbiotas.” ou “Os indivíduos que tomam antibióticos, tendem a ter uma quantidade menor de microorganismos”. Desta forma, ao apenas descrever o observado, os/as estudantes não excluíram nenhum dos quatro cenários possíveis: i. alterações em x resultam em alterações em y (x causa y); ii. alterações em y resultam em alterações em x (y causa x); iii. alterações em z resultam em alterações em x e y (z causa x e z causa y); e iv. a correlação é fruto do acaso (z causa x e w causa y), descritos por Curran-Everett (2010) e Barrowman (2014), que justificam a existência de uma correlação.

Um segundo grupo de estudantes também usou os dados observados na tabela, porém, ao explicá-los, usou termos que restringiam as relações observadas ao cenário i. alterações em x resultam em alterações em y (x causa y), estabelecendo então relações de causalidade. É exemplo disso as seguintes unidades de sentido: “A % de gordura corporal altera a quantidade de organismos na microbiota intestinal.”, “Consumo de frutas, altera a quantidade também.”, “Por fim, o consumo de antibióticos altera quais organismos vão aparecer”. Neste caso, o uso da palavra ‘altera’ traz o sentido de que tais características são as causas diretas da variação na microbiota, excluindo a possibilidade dos cenários ii, no qual a alteração na microbiota que causaria alterações na característica citada; do cenário iii, no qual alterações em uma outra variável causaria alterações na característica citada e na microbiota; e do cenário iv, no qual

DOI: 10.46667/renbio.v16i1.904

tais alterações não estariam relacionadas e seriam apenas um acaso. (CURRAN-EVERETT, 2010; BARROWMAN, 2014)

Por último, um terceiro grupo de estudantes evocou seus conhecimentos prévios para justificar os dados de correlações encontrados na tabela. Desta forma, assim como no grupo anterior, eles restringiram as relações ao cenário i de Curran-Everett (2010) e Barrowman (2014), estabelecendo relações causais entre as variáveis. São os casos de respostas como as seguintes: “O número mais disparado é o de uso de antibióticos, em que quase todos que não usaram no mês possuem consideravelmente mais gêneros presentes do que quem os usou, visto que obviamente os antibióticos matam as bactérias, sejam elas boas ou más” e “Podemos ver que o consumo de frutas pode influenciar muito no aumento de bactérias no organismo já que as frutas não possuem conservantes como as comidas industrializadas, é mais propícia a possuir bactérias.”

No Quadro 2 a seguir, mostramos o resultado da nossa análise dos diferentes tipos de relações estabelecidos pelos estudantes na atividade 2. Ressaltamos que os dados das características sexo e consumo de café não foram incluídos por não estarem correlacionadas à composição da microbiota.

**Quadro 2:** Análise dos diferentes tipos de relações estabelecidas na atividade 1.

Relação estabelecida pelos estudantes	Características				Total
	Gordura corporal	Consumo de frutas	Hábito intestinal	Consumo de antibióticos	
Correlação	3	2	3	6	14
Causalidade	2	5	3	6	16
Total	5	7	6	12	30

Fonte: autoria própria

DOI: 10.46667/renbio.v16i1.904

No Quadro 2 também podemos observar que a característica que os/as estudantes mais relacionaram com a microbiota, tanto estabelecendo correlação, quanto causalidade, foi o consumo de antibióticos, totalizando 40% das relações estabelecidas. Acreditamos que tal resultado possa estar relacionado com o fato de que o antibiótico e seus efeitos nocivos para as bactérias são assuntos amplamente discutidos tanto na mídia quanto no ambiente escolar e, portanto, os/as estudantes tendem a analisar os dados relacionados a esse tópico levando em conta os conhecimentos que já possuem sobre o tema. Isso facilita o reconhecimento da existência de uma relação entre a microbiota e o uso de antibióticos, porém, tendo em vista que a relação entre essas variáveis foi a que mais vezes foi interpretada erroneamente como causalidade, podemos pensar que o conhecimento prévio também pode induzir o/a estudante a estabelecer uma relação de causalidade entre essas variáveis, mesmo os dados demonstrando apenas uma correlação.

Também vemos isso refletido nos resultados da atividade 5, que tratava exclusivamente da relação entre o uso de antibióticos e a composição da microbiota. Tal atividade propunha que os/as estudantes escolhessem e justificassem o grau de confiança para as seguintes afirmações: “Os antibióticos matam bactérias e, por isso, há uma diminuição da diversidade da microbiota em indivíduos que utilizam esse tipo de medicação” e “Indivíduos que têm uma microbiota muito diversa têm mais chances de ter no organismo certas bactérias que protegem de doenças e, por causa dessas bactérias, necessitam de menor uso de antibióticos”.

Relativas a esta atividade, recebemos 10 respostas e escolhemos analisar apenas as que continham justificativas. Dentre as respostas, obtivemos nove justificativas no total, sendo quatro para a primeira constatação (o uso de antibióticos altera a diversidade da microbiota) e cinco para a segunda (a diversidade da microbiota altera a necessidade do uso de antibióticos). Nós as classificamos em dois grupos: justificativas baseadas em dados da tabela e justificativas baseadas em conhecimentos prévios. Apesar do baixo número de justificativas, observando o Quadro 3, vemos que a maioria dos/as estudantes que utilizou dados da tabela em sua resposta escolheu a alternativa certa. Já os que usaram dados de conhecimentos prévios acabaram, em sua maioria, escolhendo a alternativa errada, que afirmava uma relação causal entre as variáveis. Desta forma, mesmo após a videoaula sobre correlação e causalidade, os/as estudantes realizaram a mesma prática da atividade 2, ou seja, usar dados de correlação para justificar uma causalidade. Além disso, reforçando o que constatamos na atividade anterior, os alunos e as alunas que utilizaram conhecimentos prévios sobre o assunto, em sua maioria, tiveram uma interpretação enviesada dos dados.

Ponderamos que essa mobilização de conhecimentos prévios constatada nas atividades da SD também ocorre quando as pessoas interpretam manchetes e textos de divulgação científica e é nesse momento que a correlação acaba sendo interpretada como causalidade, como é mencionado no editorial da Revista Nature Medicine (2001). Ressaltamos que o problema não é a mobilização de conhecimentos prévios em si, mas sim a maneira equivocada de articular esses conhecimentos com outras informações, como os dados fornecidos pela tabela, no caso dos/as participantes da SD, ou as informações advindas de pesquisas científicas, no caso da população.

**Quadro 3:** Análise das justificativas da Atividade 4.

Constatação	Justificativa com dados fornecidos		Justificativa com conhecimento prévio	
	Resposta certa	Resposta errada	Resposta certa	Resposta errada
Antibiótico -> Microbiota	2	1	0	1
Microbiota -> Antibiótico	2	0	1	2

Fonte: autoria própria

Para finalizar, na atividade 7, três dos quatro grupos analisados escolheram a alternativa correta para a questão proposta. Ao observarmos as justificativas apresentadas pelos grupos e os diálogos ocorridos durante a execução da atividade, porém, foi possível perceber que ainda restavam muitas dúvidas e confusão, por parte dos/as estudantes, a respeito das relações de correlação e causalidade. No Quadro 4, podemos encontrar exemplos de diálogos ou de justificativas que demonstram essas confusões.

**Quadro 4:** Exemplos que demonstram as confusões a respeito dos conceitos de correlação e causalidade.

Grupo	Retirado de	Trecho
Grupo 1	Diálogo via chat	“A resposta certa é A pois <b>os dois textos são uma causalidade pois quanto mais tem a diversidade de microbiotas mais magra é a pessoa.</b> ”
Grupo 2	Diálogo	“ <b>Um acaso. Um acaso aconteceu mas não faz... pode não acontecer, tá ligado?</b> ” (sobre o que é causalidade)
Grupo 3	Diálogo	“... <b>Eu acho que é, tipo, tem a ver sabe? A microbiota tem a ver, você pode ter certeza que tem a ver com o câncer sei lá o que.</b> ” (sobre o que é correlação)
Grupo 4	Justificativa via chat	“... <b>Já no segundo texto, é apresentado que o microbioma pode ou não ter relação com a obesidade, sendo assim uma casualidade.</b> Devido a fatos e argumentos acreditamos que possa ser equivocado tal afirmação, ou até mesmo ter errado em sua tese.”

Fonte: autoria própria

No trecho extraído do grupo 3, podemos ver que, ao tentar explicar o que é uma correlação, o/a estudante exclui de sua explicação a possibilidade da ocorrência do cenário iv de Curran-Everett (2010) e Barrowman (2014), que diz que uma correlação pode ser fruto de um acaso. Já nos trechos dos grupos 2 e 4, vemos que os/as estudantes não entenderam que uma causa deve ser consistente em seu efeito em condições constantes (KUHN et al., 2015). Por último, no trecho do grupo 1, ao explicar o motivo da existência de uma causalidade entre duas variáveis, o/a estudante deixa em aberto a possível existência dos quatro cenários descritos por Curran-Everett (2010) e Barrowman (2014), portanto, o que ele descreve é a existência de uma correlação e não causalidade. Esses dados constituem-se como evidência de que estudantes enfrentam dificuldades para perceber a diferença entre correlação e causalidade – mesmo após terem sido expostos a uma sequência didática desenvolvida para abordar especificamente esses conceitos.

Conforme afirmamos anteriormente, para estruturar a SD utilizada nesta pesquisa, utilizamos o Ensino de Ciências por Investigação, uma abordagem didática na qual os estudantes i. são envolvidos em questões cientificamente orientadas; ii. dão prioridade às evidências para abordar questões cientificamente orientadas e propor explicações; iii. avaliam suas explicações à luz de proposições alternativas; e iv. comunicam e justificam as explicações propostas (NATIONAL RESEARCH COUNCIL, 2000).

As atividades propostas na SD, como vimos, ofereceram oportunidades para os estudantes se engajarem na maior parte das práticas científicas elencadas acima, mas avaliamos que a modalidade de ensino remoto inviabilizou uma dimensão importante da abordagem investigativa: a interação com os pares. Carvalho (2013) afirma que o EnCI deve ser planejado considerando a importância da socialização na construção do conhecimento. Scarpa e Campos (2018) também apontam que a interação e a colaboração entre as/os estudantes são aspectos essenciais dessa abordagem didática.

No contexto observado por meio desta pesquisa, as atividades de análise de dados (atividade 2 no 1º ciclo investigativo e atividade 5 no 2º ciclo investigativo) foram realizadas de maneira assíncrona e individualmente. Dessa forma, a interação social entre os estudantes, a professora, os problemas de investigação, os dados, os valores epistêmicos e sociais para a construção do conhecimento não foi privilegiada. Sem a interação com a professora e os/as colegas, não é dada a possibilidade de o erro ser reconhecido, trabalhado e superado pelos/as próprios/as alunos/as.

Dada a complexidade das relações entre variáveis, acreditamos que a distinção entre correlação e causalidade precisa ser desenvolvida ao longo de toda a escolaridade dos alunos e das alunas, uma vez que a compreensão limitada de tais relações interfere na habilidade de interpretação de dados por parte dos discentes, o que também atrapalha o entendimento de outros conceitos científicos e a tomada de decisão embasada em questões cotidianas relacionadas à ciência.

#### 4 Considerações finais

Em sua fala durante a conferência de encerramento do 2º Encontro de Ensino de Ciências por Investigação (ENECCI), ocorrido na UFMG na modalidade remota em outubro de 2020, Martins ressaltou a importância de as pessoas conhecerem o processo de produção do conhecimento científico, bem como suas formas de validação e legitimação, para que o seu caráter provisório não seja confundido com imprecisão e, desta forma, possamos contribuir para que *fake news* não sejam disseminadas. Ainda segundo ela, a capacidade de distinguir correlações de relações causais é um importante meio para se atingir tal objetivo (CECIMIG FaE UFMG, 2020).

Ao nosso ver, isso se justifica porque, apesar de os estudos baseados em relações probabilísticas (correlação) serem de grande valor para a comunidade científica (por justificarem a importância de se realizar novas pesquisas sobre os temas estudados e suas relações, por exemplo), frequentemente tais estudos são divulgados por diversos meios de comunicação em massa (mídia corporativa, alternativa, redes sociais etc.) e a correlação demonstrada acaba sendo interpretada como uma causalidade pela população, que tende a desacreditar da ciência se tal relação de causalidade não for comprovada em pesquisas futuras.

Recentemente, durante a pandemia do COVID-19, pudemos ver exemplos disso, como alguns estudos observacionais que mostraram uma correlação entre o uso de hidroxicloroquina e a cura de pacientes com COVID-19. Tais estudos, ao serem divulgados para a população, foram usados como evidência de uma relação causal entre o uso desse medicamento e a cura da COVID-19, gerando uma grande confusão sobre o assunto, mesmo depois de outros estudos negarem a existência dessa relação causal.

Tendo isso em vista, tratar sobre correlação e causalidade na escola é importante para a formação de pessoas mais críticas, que sejam capazes de articular seus conhecimentos para a tomada de decisões. A SD desenvolvida neste estudo trouxe o tema da microbiota intestinal para abordar estes conceitos, uma vez que são vários os fatores que estão relacionados com a variação na composição da microbiota intestinal, mas a correta compreensão dos dois conceitos é fundamental não apenas para a compreensão de aspectos próprios das ciências, mas também para as mais diversas esferas da vida das pessoas.

Com a aplicação da SD, percebemos uma grande dificuldade dos/as estudantes de distinguirem correlação de causalidade. Durante todas as atividades, alunos/as usaram dados que demonstravam correlação para justificar causalidades e, além disso, percebemos que os conhecimentos prévios dos/as estudantes, advindos do senso comum, acabaram por induzir suas interpretações sobre os dados a serem analisados, levando-os ao erro. Desta forma, é preciso atuar favorecendo a compreensão sobre a existência e o significado das correlações e ao mesmo tempo cuidar da forma de expressar relações entre variáveis.

A abordagem do Ensino de Ciências por Investigação, ao promover a análise de diferentes tipos de dados na resolução de uma questão cientificamente orientada, mostra-se propícia para a constituição de um ambiente de aprendizagem no qual discussões sobre as concepções de correlação e causalidade possam ser feitas de maneira contextualizada. Considerando que a interação e a colaboração entre as/os estudantes são aspectos essenciais dessa abordagem didática, os momentos de discussão e reflexão entre os pares podem ser fundamentais para desenvolver a compreensão dos modelos de correlação e causalidade por parte dos/as estudantes.

Ademais, sugerimos a ampliação das pesquisas acerca de como podemos utilizar os conhecimentos prévios dos/as estudantes de forma a contribuir para uma correta interpretação de dados sobre correlação e causalidade.

## Referências

ALLCHIN, Douglas. **Teaching the Nature of Science: Perspectives & Resources**. St. Paul, MN: SHiPS Education Press, 2013.

BARROWMAN, Nick. Correlation, causation, and confusion. **The New Atlantis**, p. 23-44, 2014.

BRAATEN, Melissa; WINDSCHITL, Mark. Working toward a stronger conceptualization of scientific explanation for science education. **Science education**, v. 95, n. 4, p. 639-669, 2011.

CARVALHO, Anna Maria Pessoa de et al. O ensino de ciências e a proposição de sequências de ensino investigativas. **Ensino de ciências por investigação: condições para implementação em sala de aula**. São Paulo: Cengage Learning, v. 1, p. 1-19, 2013.

CECIMIG FaE UFMG. **Conferência de encerramento do 2º Encontro de Ensino de Ciências por Investigação**. YouTube, 16/10/2020. Disponível em: <https://www.youtube.com/watch?v=2wiWwFw9upM>. Acesso em: 25 abr. 2023

CHINN, Clark A.; MALHOTRA, Betina A. Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. **Science education**, v. 86, n. 2, p. 175-218, 2002.

CURRAN-EVERETT, Douglas. Explorations in statistics: power. **Advances in Physiology Education**, v. 34, n. 2, p. 41-43, 2010.

DEVROOP, Karendra. **Correlation versus Causation: Another Look at a Common Misinterpretation**. 2000.

DUSCHL, Richard A.; GRANDY, Richard E. Reconsidering the character and role of inquiry in school science: Framing the debates. **Teaching scientific inquiry**. Leiden, The Netherlands: Brill, 2008. p. 1-37.

EDITORIAL. In science we trust. **Nature Medicine**, v. 7, p. 871, 2001.

FALONY, Gwen et al. Population-level analysis of gut microbiome variation. **Science**, v. 352, n. 6285, p. 560-564, 2016.

HALL, Kevin D. et al. Ultra-processed diets cause excess calorie intake and weight gain: an inpatient randomized controlled trial of ad libitum food intake. **Cell metabolism**, v. 30, n. 1, p. 67-77. e3, 2019.

KELLY, Gregory J.; LICONA, Peter. Epistemic practices and science education. **History, philosophy and science teaching**. Springer, Cham, p. 139-165, 2018.

DOI: 10.46667/renbio.v16i1.904

- KELLY, Gregory. Inquiry, activity and epistemic practice. **Teaching scientific inquiry**. Leiden, The Netherlands: Brill, p. 99-117, 2008.
- KUHN, Thomas S. **A estrutura das revoluções científicas**. São Paulo: Editora Perspectiva, 2020.
- KUHN, Deanna; RAMSEY, Stephanie; ARVIDSSON, Toi Sin. Developing multivariable thinkers. **Cognitive Development**, v. 35, p. 92-110, 2015.
- LONGINO, Helen E. **The fate of knowledge**. Princeton: Princeton University Press, 2018.
- MCCAIN, Kevin. Explanation and the nature of scientific knowledge. **Science & Education**, v. 24, n. 7, p. 827-854, 2015.
- MICHAEL, Joel. What makes physiology hard for students to learn? Results of a faculty survey. **Advances in physiology education**, v. 31, n. 1, p. 34-40, 2007.
- MINNER, Daphne D.; LEVY, Abigail Jurist; CENTURY, Jeanne. Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. **Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching**, v. 47, n. 4, p. 474-496, 2010.
- MORAES, Roque; GALIAZZI, Maria do Carmo. Análise Textual Discursiva: processo constitutivo de múltiplas faces. **Ciência & Educação**. Bauru, v.12, n.1, p. 117-128, 2006.
- NATIONAL RESEARCH COUNCIL. **Inquiry and the National Science Education Standards: A Guide for Teaching and Learning**. Washington, DC: The National Academies Press, 2000.
- NORRIS, Stephen P. et al. A theoretical framework for narrative explanation in science. **Science Education**, v. 89, n. 4, p. 535-563, 2005.
- OSBORNE, Jonathan. Defining a Knowledge Base for Reasoning in Science: The Role of Procedural and Epistemic Knowledge. **Reconceptualizing STEM Education**. Routledge, p. 229-245, 2016.
- PEDASTE, Margus et al. Phases of inquiry-based learning: Definitions and the inquiry cycle. **Educational research review**, v. 14, p. 47-61, 2015.
- SCARPA, Daniela Lopes; CAMPOS, Natália Ferreira. Potencialidades do ensino de Biologia por Investigação. **Estudos avançados**, v. 32, p. 25-41, 2018.
- SLOMINSKI, Tara; GRINDBERG, Sarah; MOMSEN, Jennifer. Physiology is hard: a replication study of students' perceived learning difficulties. **Advances in physiology education**, v. 43, n. 2, p. 121-127, 2019.

DOI: 10.46667/renbio.v16i1.904

STREVENS, Michael. No understanding without explanation. **Studies in history and philosophy of science Part A**, v. 44, n. 3, p. 510-515, 2013.

ZORZETTO, Ricardo. Alguns efeitos dos alimentos fabricados. **Revista Pesquisa FAPESP**, São Paulo, n. 265, p. 21-27, 2018.

Recebido em janeiro de 2023.  
Aprovado em maio de 2023.

Revisão gramatical realizada por: Valéria Melki Busin  
E-mail: [valeriamelkibusin@gmail.com](mailto:valeriamelkibusin@gmail.com)