



CONCEPÇÕES DE ALUNOS DO ENSINO MÉDIO SOBRE TAMANHO DA AMOSTRA E REPRODUTIBILIDADE DE DADOS EXPERIMENTAIS

High school students' understanding of sample size and reproducibility in experimental data

Alessandro Damásio Trani Gomes [alessandro.gomes@ufsj.edu.br]

Departamento de Ciências Naturais

Universidade Federal de São João del-Rei

Praça Dom Helvécio, 74, São João del-Rei, Minas Gerais, Brasil

Resumo

Este trabalho investiga as concepções de alunos do ensino médio sobre os conceitos de evidência relativos à reprodutibilidade de dados experimentais. Avalia também a evolução desses conceitos ao longo de um ano letivo, considerando a realização de uma sequência de atividades práticas diversas, integradas ao currículo da escola. Ao todo, 164 alunos do 1º ano do ensino médio participaram integralmente da pesquisa, com base em uma abordagem quantitativa, com delineamento pré-experimental, consistindo na aplicação de dois testes, um no início e outro no final do ano letivo. Além dos testes, alguns alunos foram entrevistados. A comparação dos resultados do pré e do pós-teste indica, para o grupo de alunos pesquisado, que houve certa sofisticação no pensamento acerca dos construtos pesquisados, porém, ao se analisarem as justificativas, percebe-se que grande parte dos participantes não compreendeu adequadamente os motivos reais da importância de se trabalhar com um número maior de medidas. Com base nos resultados obtidos, são discutidas as implicações educacionais e propostas novas questões de pesquisa.

Palavras-Chave: Medição; Análise de dados; Atividades práticas; Reprodutibilidade; Laboratório.

Abstract

This study investigates the perceptions of high school students regarding the concepts of evidence related to the reproducibility of experimental data, also in addition to analyzing the evolution of these concepts over a school year, considering the implementation of a series of diverse practical activities integrated into the school curriculum. In total, 164 first-year high school students fully participated in the research, based on a quantitative approach covering a pre-experimental design with two tests, one at the beginning and another at the end of the school year. In addition to the tests, some students were interviewed. The comparison of the pre- and post-test results indicates, for the group of students surveyed, that there was a certain sophistication in thinking about the constructs researched. However, the analysis of the justifications revealed that a large portion of the participants did not understand the real reasons of the importance of working with a larger number of measurements properly. Based on the results, educational implications are discussed, and new research questions are proposed.

Keywords: Measurement; Data analysis; Practical activities; Reproducibility; Laboratory.

INTRODUÇÃO

Para fazer face às exigências formativas do século XXI e se tornarem cidadãos cada vez mais críticos, os alunos precisam ser capazes de utilizar conhecimentos científicos para contribuir com o debate público e desenvolver opiniões informadas sobre questões fundamentadas na ciência e na tecnologia (OCDE, 2019). Subjacente a esse importante objetivo está a noção de alfabetização científica (Gräber, Nentwig, Koballa, & Evans, 2013; Sasseron & Carvalho, 2011), a qual proporciona condições para que as pessoas possam abordar, tomar decisões e atuar sobre problemas cotidianas e complexas, mobilizando estratégias, valores, habilidades, competências e conhecimentos necessários.

Envolver os alunos em práticas que, em certos aspectos, se assemelham às atividades desenvolvidas pelos cientistas é uma das formas de se promover a alfabetização científica. Consequentemente, educadores e currículos em todo o mundo procuram promover, sistematicamente, o conhecimento e as habilidades necessárias para que os alunos as compreendam e as realizem de forma satisfatória (NRC, 2012; NGSS Lead States, 2013; KMK, 2020). Essas práticas abrangem

“uma série de atividades que incluem formular questões, desenvolver hipóteses, projetar e conduzir experimentos, examinar e interpretar dados, construir argumentos e contra-argumentos e debater conclusões. Acredita-se que será mais eficaz se os alunos se envolverem nestas atividades em primeira mão, não como procedimentos isolados, mas como aspectos interligados de uma investigação ampla baseada em objetivos que abordam uma ou mais questões significativas” (Kuhn, Arvidsson, Lesperance, & Corprew, 2017, p. 233).

Compreender a ciência como prática, seu funcionamento e desenvolvimento requer conhecimentos conceituais, procedimentais e epistêmicos (Osborne, 2016). O conhecimento conceitual consiste no entendimento das ideias da ciência, as quais são baseadas em fatos, leis. O conhecimento epistêmico se refere à compreensão do papel de características e construtos essenciais e definidores para o processo de construção do conhecimento científico, incluindo o entendimento sobre a função que perguntas, hipóteses, teorias, modelos e argumentos desempenham na atividade científica (OCDE, 2019). O conhecimento procedimental, antes concebido como sendo composto por técnicas práticas, destrezas manuais, procedimentos padrões e habilidades operacionais gerais, passou a ser considerado, nas últimas décadas, como composto e influenciado por conceitos e ideais, sobretudo relacionadas à produção e avaliação de evidências (Millar, Lubben, Got, & Duggan, 1994; Gott & Duggan, 1995).

Em seu trabalho, Gott e Duggan (1995) utilizam o termo ‘conceitos de evidência’ para se referirem aos conceitos associados ao conhecimento procedimental que dizem respeito aos raciocínios, às ideias e às habilidades necessárias para, além de se coletar dados confiáveis e válidos, organizá-los, tratá-los e interpretá-los apropriadamente para serem utilizados na avaliação de evidências, hipóteses e explicações. Dentre os conceitos de evidência elencados por Gott e Duggan (1995) estão os relacionados ao tamanho da amostra e a reprodutibilidade dos dados experimentais, ou seja, a ideia de que qualquer processo de medição, quando repetido sistematicamente, produz uma série de valores, bem como que para se analisarem tais resultados, é necessária uma compreensão sobre a dispersão dos dados obtidos e as causas dessa variabilidade.

Em uma recente revisão da literatura, Pigosso e Heidemann (2023) afirmam que a pesquisa sobre o processo de medição científica no ensino de ciências ainda é incipiente no Brasil, tendo sido contabilizados apenas 8 artigos nos últimos 30 anos. Também são escassos os instrumentos de pesquisa que avaliam a compreensão de alunos a respeito desse processo (Pols, Dekkers, & de Vries, 2022). Sendo assim, este trabalho busca preencher essa lacuna ao buscar respostas para duas questões:

1) Em que medida alunos ingressantes do ensino médio compreendem os conceitos relativos ao tamanho da amostra e à reprodutibilidade de dados experimentais?

2) O domínio de conceitos relativos ao tamanho da amostra e à reprodutibilidade de dados experimentais aumenta ao longo de um ano letivo, após a realização de uma sequência de atividades práticas diversas?

As ações de identificar e analisar as ideias e dificuldades dos alunos não apenas tornam explícitas suas concepções implícitas, mas também permite abordá-las diretamente para desafiá-las e transformá-las, reconhecendo-as como pontos de partida importantes para a aprendizagem. Isso pode ser alcançado ao se utilizarem os resultados de pesquisa como uma fonte potencial de aprendizagem: as dificuldades dos alunos e os padrões de respostas obtidas têm o potencial de servir como orientações para o planejamento de ações pedagógicas e novas pesquisas na área (Kranz, Baur, & Mölle, 2023).

O TAMANHO DA AMOSTRA E A REPRODUTIBILIDADE DOS DADOS EXPERIMENTAIS

Segundo educadores e elaboradores de currículos, os conceitos envolvidos no processo de medição científica são fundamentais para a compreensão do empreendimento científico e devem ser ensinados desde os anos iniciais da educação básica (NRC, 2012; Camargo Filho, Laburú, & Barros, 2015; MEC, 2018).

O número de medidas diferentes realizadas em um experimento está relacionado diretamente ao tamanho da amostra e ao intervalo entre as medidas em um conjunto de dados. Supondo que deseje determinar a relação entre duas variáveis contínuas (que podem ser representadas por qualquer número real), um aluno precisa modificar a variável independente para produzir um certo número de valores distintos na variável dependente, de forma que esses valores sejam separados por um intervalo apropriado e distribuídos de forma mais ou menos homogênea dentro dos limites de medição estabelecidos. Caso isso não ocorra, a relação entre as variáveis poderá não ser corretamente identificada. Pesquisas indicam que grande parte dos alunos, em todos os níveis de ensino, não dão a devida atenção a questões relativas ao número de dados a serem coletados, aos limites e aos intervalos de medição ao planejarem e executarem atividades práticas (Gott & Duggan, 1995).

“Estas práticas de medição e a aprendizagem pretendida do conceito de incerteza são muitas vezes difíceis de compreender. Como resultado, as tarefas rotineiras de formação nos cursos laboratoriais tradicionais reforçam a crença dos alunos na existência de um valor verdadeiro; as incertezas nas medições são vistas como erros, mas não como propriedade intrínseca de todas as medições. Em outras palavras, eles geralmente acreditam que, a princípio, pode ser feita uma medição perfeita sem qualquer incerteza. Além disso, não compreendem, normalmente, a necessidade de fazer medições repetidas e muitas vezes têm uma noção geral de que medições repetidas produzem um resultado melhor, sem compreenderem o que realmente significa “melhor resultado”. Na operação, os alunos sempre tratam a média aritmética como o resultado final de um conjunto de dados, que é tudo o que importa na comparação de dois conjuntos de dados. Estes mal-entendidos muitas vezes os tornam incapazes de distinguir entre incertezas de origem aleatória e sistemática e não conseguem identificar diferentes fontes de incerteza numa medição. Todas essas dificuldades sugerem que é um desafio para os alunos alcançar uma compreensão adequada e desenvolver uma estrutura de conhecimento integrada ao aprender sobre o processo de medição” (Lu et al., 2023, p.2).

Schauble (1996) examinou como alunos entre 10 e 13 anos lidam com a variabilidade dos valores obtidos resultantes de erros experimentais e erros de medida. As crianças trabalharam em dois sistemas envolvendo líquidos e objetos submersos com o objetivo de determinar a influência das variáveis envolvidas. Como resultado, a autora aponta que as crianças participantes não mencionaram nos seus planos de investigação a necessidade de repetição de medidas. Elas acreditavam que a medida inicial é a ‘correta’. Ao repetirem as medidas, as crianças demonstravam ficar confusas e surpresas até mesmo com pequenas diferenças entre os valores. Schauble ressalta a dificuldade das crianças de distinguir variações de medidas devido a erros experimentais das variações provocadas devido às mudanças em determinada variável causal.

De mesmo modo, Varelas (1997) descreve que alunos entre 9 e 10 anos demonstram a visão de que para uma mesma condição experimental, apenas um resultado para as medidas seria possível. Alunos que reconhecem que pequenas variações podem ocorrer (sob condições experimentais equivalentes) não ficam tão preocupados e confusos com a obtenção de valores distintos para uma mesma medida, em comparação àqueles que possuem uma visão mais ingênua do processo de medição.

Para alunos com idades compatíveis àquelas do ensino médio, a situação não é tão melhor. Muitos reconhecem a necessidade de se repetirem as medidas, mas poucos conseguem articular uma razão para tal procedimento (Gott & Duggan, 1995; Lubben & Millar, 1996). A situação é semelhante com alunos no ensino superior (Séré, Journeaux, & Larcher, 1993; Allie, Buffler, Campbell, & Lubben, 1998). Muitos possuem a concepção de que repetir medidas é apenas uma atividade de rotina para que se possa efetuar o cálculo da média dos valores medidos.

Séré e colaboradores (1993) trabalharam com universitários em um curso de física, no qual o objetivo da sequência de ensino era ensinar e trabalhar os conceitos envolvidos no processo de medição. Vinte alunos participaram da pesquisa, realizando atividades relacionadas a temas de óptica e circuitos elétricos. Os autores observaram o comportamento dos participantes durante o laboratório, analisaram seus relatórios e realizaram um exame final. Além disso, entrevistaram 4 desses alunos com o objetivo de identificar sua compreensão sobre os conceitos trabalhados. Os autores perceberam que, em geral, os alunos demonstraram proficiência em aplicar certos procedimentos algorítmicos, mas pouca compreensão sobre o que realmente faziam e o porquê. Eles repetiam medidas, sobretudo, quando acreditavam que tinham motivos para descartar a primeira medida. Mesmo realizando várias medidas, os participantes

preferiam, normalmente, o primeiro dado coletado ou o valor mais recorrente como o valor mais representativo do conjunto dos dados. Os autores puderam identificar também que cerca da metade dos alunos:

- (i) compreendia que a realização de mais medidas conduz a melhores resultados, mas não tinham certeza do que isso significava;
- (ii) colocava maior confiança no primeiro resultado ou no resultado que mais se repetia ao avaliarem uma série de dados coletados;
- (iii) reportava apenas a precisão do instrumento quando solicitados a reportarem a dispersão dos dados e
- (iv) considerava como um “resultado ruim” aquele que possuía amplo desvio padrão, mas não consideravam as fontes de erros, sobretudo os erros sistemáticos.

Lubben e Millar (1996) realizaram uma pesquisa e identificaram as concepções de mais de mil alunos entre 11 e 15 anos sobre o processo de medição. O teste versava principalmente sobre a validade e reprodutibilidade de dados obtidos por meio de medidas. Dentre as 6 questões que compunham o teste, havia uma sobre a razão de se repetirem ou não as medidas, formas de lidar com resultados muito diferentes dos demais e o significado da variação dos resultados. As respostas a cada pergunta foram analisadas utilizando um sistema de codificação desenvolvido a partir do conjunto de respostas disponíveis. Como resultado da pesquisa, Lubben e Millar propõem uma série de níveis hierárquicos para a compreensão dos alunos sobre o processo de medição (Quadro 1). Cada nível é especificado por três ideias básicas: a visão sobre o processo de medição, o modo para se avaliar o resultado obtido e o método para se identificar e tratar os resultados anômalos. A progressão da compreensão parte do não reconhecimento da necessidade de se realizarem outras medidas, passando pela procura por resultados recorrentes e uma variação deliberada no controle de variáveis para garantir a variação nos resultados, alcançando então a compreensão de se determinarem a variação e a dispersão dos dados, aspectos que inicialmente aparecem mais como uma rotina das aulas experimentais, sem a devida compreensão. As respostas obtidas fornecem indícios da existência dos 8 níveis identificados pelos autores, tornando o modelo uma ferramenta importante para classificar as ações e ideias dos indivíduos sobre a reprodutibilidade dos dados. Porém, Lubben e Millar enfatizam que não obtiveram informações sobre como os indivíduos progredem entre esses níveis. Enquanto se pode esperar uma progressão natural com a idade e escolarização entre os níveis, não se acredita que o indivíduo progrida de forma linear e homogênea, podendo ‘saltar’ níveis ou estar localizado em dois níveis ao mesmo tempo.

Coelho e Séré (1998) entrevistam 21 alunos franceses de ensino médio com idades entre 14 e 17 anos durante a realização de atividades práticas envolvendo coleta de dados, nas quais os participantes trabalharam em duplas. As autoras afirmam que a maioria deles acredita na existência de um ‘valor verdadeiro’ ou um ‘valor correto’ para a medida e demonstra insatisfação ou não compreende bem as inconsistências e variações de suas medidas. Os participantes fizeram frequentemente alusão à qualidade dos equipamentos e à falta de cuidados tomados durante a medição para justificar as diferenças obtidas. Segundo as autoras, isso justifica o mito criado por alguns alunos de que os físicos e cientistas profissionais contam com técnicas e instrumentos perfeitos que possibilitam a obtenção de um valor ideal para a medida.

Fred Lubben e colaboradores realizaram uma série de trabalhos que forneceu um importante suporte empírico para o desenvolvimento de um referencial no qual o pensamento e o raciocínio utilizados pelos alunos durante o processo de medição e análise dos dados podem ser categorizados em dois tipos distintos, segundo os paradigmas Pontual e de Conjunto (Allie *et al.*, 1998; Buffler, Allie, & Lubben, 2001; Lubben, Campbell, Buffler, & Allie, 2001). O Quadro 2 mostra as diferenças entre os paradigmas.

O paradigma pontual se caracteriza pela noção de que cada medida pode, em princípio, ser o valor correto. Como consequência, cada medida é independente das outras e as medidas individuais não são combinadas de forma alguma com as outras. Portanto, cada medida é concebida como sendo um valor, em vez de contribuir para se estabelecer um intervalo. Na forma mais extrema do pensamento, segundo esse paradigma, está a crença de que apenas uma medida seria necessária para se estabelecer o valor verdadeiro da grandeza medida. O paradigma de conjunto é caracterizado pela ideia de que cada medida seria apenas uma aproximação do valor verdadeiro e que o desvio desse valor seria aleatório. Como consequência, um número de medidas é necessário para formar uma distribuição que se estabeleça em torno de um valor determinado. A melhor informação sobre o valor verdadeiro é obtida pela combinação das medidas, utilizando-se construtos teóricos para descrever os dados conjuntamente.

Quadro 1 – Níveis de compreensão do processo de medição segundo Lubben e Millar (1996)

Níveis	Visão do processo de medição	Como avaliar o resultado	O que fazer com os resultados que diferem significativamente dos demais
A	O processo de medição é direto: mede-se uma vez e obtém-se o valor correto.	Não é uma questão. As medidas são corretas.	Não é uma questão.
B	Mede-se apenas uma vez e o resultado é sua medida correta. Repetir medidas apenas trará confusão.	A menos que algo óbvio tenha saído errado, a medida que foi feita está correta. Em contextos familiares, o resultado é aquele esperado.	Não é uma questão.
C	Se utilizar equipamentos sofisticados e adequados, suas medidas serão certas. Faça algumas medidas iniciais e quando estiver familiarizado com o equipamento, faça a medida correta.	A menos que algo óbvio tenha saído errado, a medida que foi feita após algum treinamento está correta. Em contextos familiares, o resultado é aquele esperado.	Ignora-se (as diferenças que ocorreram foram devido à falta de prática).
D	Se utilizar equipamentos sofisticados e adequados, suas medidas serão certas. Devem-se repetir medidas até encontrar o mesmo valor.	Obter dois dados iguais significa que as medidas foram feitas com cuidado suficiente.	Ignora-se.
E	Devem-se repetir as medidas e calcular a média. Mas repetir as medidas da mesma forma resultará nos mesmos resultados. Devem-se variar um pouco as condições.	Variações são esperadas. Não são problemas.	Variações são esperadas. Devem-se incluir todos os dados e calcular a média.
F	Medidas cuidadosas podem se aproximar do valor correto da grandeza medida, mas nunca se saberá qual o valor correto. Tira-se a média, que representará tal valor.	Não pode ser avaliado 'em si mesmo'. O único método é checar com alguma fonte de autoridade (professor, livro etc.).	Incluem-se os dados anômalos no cálculo da média, pois tira-se a média para levar em consideração essas variações.
G	Como o nível anterior.	Pode ser avaliado 'por dentro'. O intervalo de dados é uma indicação.	Como o nível anterior.
H	Como o nível anterior.	Como o nível anterior.	É apropriado realizar uma avaliação do conjunto de dados e rejeitar dados anômalos antes de se calcular a média. A média de um conjunto de dados pode ser, portanto, melhor que a média de outro conjunto.

Fonte: traduzido de Lubben & Millar (1996, p.966-967).

Quadro 2 – Caracterização dos paradigmas pontual e de conjunto.

Paradigma pontual	Paradigma de conjunto
O processo de medição permite determinar o valor verdadeiro da grandeza.	O processo de medição fornece informações incompletas sobre a grandeza que está sendo medida.
“Erros” acontecem devido a uma inabilidade do experimentador ou devido a falhas na configuração experimental.	Todas as medidas estão sujeitas a incertezas que não podem ser eliminadas. As incertezas podem apenas ser reduzidas.
Uma única medida tem o potencial de ser o valor verdadeiro da grandeza que está sendo medida.	As medidas realizadas são utilizadas para construir uma distribuição a partir da qual a melhor aproximação da grandeza medida e a dispersão dos dados serão obtidas.

Fonte: traduzido de Lubben, Allie & Buffler (2010, p.137).

O referencial proposto por Buffler *et al.* (2001) é importante porque define claramente o objetivo da instrução em relação ao processo de medição científica: estimular e provocar nos alunos a mudança progressiva do paradigma pontual para o de conjunto, o que, para Laburú e Barros (2009, 158-159) depende

“de uma superação cognitiva. Uma superação que só acontece se houver uma condução da concepção de medição, estruturada na pseudo-necessidade de um particular e obrigatório valor único e exato, que prescinde de uma noção de um intervalo de valores, em direção ao conceito de um provável “valor alvo” desconhecido, intrinsecamente correlacionado a também uma provável incerteza, igualmente desconhecida, ficando ambos vinculados, em razão disso, à esfera cognitiva das possibilidades”.

Camargo Filho *et al.* (2015) fazem um refinamento teórico para caracterizar os paradigmas propostos por Buffler *et al.* (2001) em termos de categorias de compreensão conceitual, resultando em uma análise mais detalhada e qualitativa das concepções dos alunos relativas ao processo de medição. Os autores propõem 5 níveis de compreensão conceitual de medição, de concepções alternativas diversas até a compreensão cientificamente adequada. Após a realização de um curso com alunos de física e áreas afins, os autores advertem que mesmo que os alunos tenham evoluído na direção do paradigma de conjunto, a maioria deles ainda manteve fragmentos de concepções alternativas em seus discursos.

ASPECTOS METODOLÓGICOS

Para que fosse possível responder às questões de pesquisa propostas, realizou-se uma pesquisa longitudinal de natureza quantitativa com delineamento pré-experimental, ou seja, envolveu apenas um grupo, que foi submetido a um pré-teste, seguido de tratamento e um pós-teste (Moreira, 2011). Em um estudo longitudinal, devem-se ter duas preocupações básicas. A primeira é a adoção de um espaçamento adequado entre as coletas de dados para permitir que as mudanças e os conhecimentos adquiridos pelos indivíduos possam ser assimilados e incorporados. A segunda é a utilização de instrumentos semelhantes, passíveis de comparação entre si nas coletas de dados que podem ocorrer (Gaya & Bruel, 2019). Essa estratégia é comumente utilizada em pesquisas sobre as concepções de alunos de todos os níveis de ensino sobre o processo de medição (Buffler *et al.*, 2001; Pollard, Werth, Hobbs, & Lewandowski, 2020, por exemplo).

Participantes

A coleta de dados ocorreu em uma escola da rede federal de ensino de Belo Horizonte. Dos 261 alunos matriculados no 1º ano do ensino médio, 164 participantes responderam aos dois testes aplicados. Os participantes foram divididos em 8 turmas incluindo de 30 a 34 alunos cada. Quatro professores de física eram responsáveis, cada um, por duas turmas.

Conforme os preceitos éticos recomendados pela Associação Nacional de Pós-Graduação e Pesquisa em Educação (ANPEd, 2019), a pesquisa foi devidamente autorizada pelo diretor da escola e pelos professores que auxiliariam no processo de coleta de dados. O Termo de Consentimento Livre e Esclarecido (TCLE) foi enviado aos pais e/ou responsáveis de todos os participantes. Posteriormente, os alunos entregaram o canhoto do TCLE assinado. A pesquisa também foi detalhada aos participantes antes do processo de coleta de dados. Foram esclarecidos detalhes sobre a justificativa, os objetivos, os métodos, os possíveis riscos e benefícios da pesquisa e, em seguida, colheu-se o assentimento dos alunos, sendo que todos concordaram em participar.

Nessa escola, o currículo de física tinha, à época, uma componente experimental, aproveitando-se da estrutura de laboratórios e os recursos disponíveis e os alunos realizavam atividades práticas quinzenalmente. Ao longo do ano letivo, foram desenvolvidas 13 atividades no laboratório de física, todas integradas ao currículo normal de física do 1º ano do ensino médio da escola. As atividades envolveram conceitos de mecânica (movimento uniforme e variado, energia, sistema massa-mola e força elástica) e eletricidade (circuitos simples com pilhas e lâmpadas e medição de corrente e tensão em circuitos série e paralelo). As atividades possuíam diversos níveis de abertura e especificação (Borges, 2002) e tinham diferentes objetivos, dentre os quais: aprender a formular e explicitar suas hipóteses, aprender como comunicar os resultados da atividade (por meio de relatórios), aprender como processar, representar e analisar dados, além de trabalhar os conteúdos específicos de cada atividade.

Ressalta-se que os alunos tiveram, no início do ano, três aulas teóricas (50 minutos cada) sobre teoria de erros, tipologia de erros e cálculo do valor médio e desvio absoluto médio. Além disso, a primeira das 13 atividades foi especialmente planejada para trabalhar diretamente esses conceitos relativos ao processo de medição. Porém, nas demais atividades práticas ao longo do ano, aspectos relacionados ao processo de medição, como a reprodutibilidade e a dispersão dos dados, não foram abordados de forma explícita.

Instrumento e procedimentos de pesquisa

Um fator importante durante a realização de uma atividade experimental é decidir sobre a quantidade de dados que devem ser coletados e o que será feito com o conjunto de dados. Além do tamanho da amostra, alguns conhecimentos fundamentais são reflexos das escolhas dos alunos: a noção da irreprodutibilidade dos fenômenos físicos e a consciência dos erros durante a experimentação. Para se

avaliarem esses conhecimentos, utilizou-se um instrumento de pesquisa (Figura 1) composto por 4 questões que têm como contexto um domínio familiar aos alunos participantes já utilizado em pesquisas anteriores (Allie *et al.*, 1998).

As questões são contextualizadas sobre uma atividade fictícia realizada por grupos de alunos durante uma aula no laboratório de física. Na questão, um grupo de três estudantes discute sobre as estratégias para coletarem os dados para alcançar o objetivo proposto. Eles decidem que coletarão os dados para 5 alturas h diferentes (10, 15, 20, 25, 30 cm). Porém, discordam sobre as medições da distância d para cada altura. É apresentado um comentário de cada estudante do grupo. Eles divergiam principalmente sobre a quantidade de medidas a serem realizadas para cada altura h . O aluno deve então escolher aquela que considera mais adequada e apresentar uma justificativa.

A questão 2 descreve que um grupo de estudantes começa a atividade e solta a esfera de uma altura h de 20 cm. Eles medem a distância d alcançada pela esfera ao atingir o chão, obtendo o valor de 67,4 cm. O grupo de estudantes resolve realizar nova coleta de dados com a mesma altura. Então, é perguntado ao respondente qual o valor que ele esperaria que o grupo obtivesse. Também é solicitado que se forneça uma justificativa para a resposta. Nesse item, explorou-se ainda mais a noção de erros de medida. Espera-se que os alunos que acreditam não haver mínimas variações de um experimento para outro, bem como que possíveis erros não possam acontecer, respondam que o grupo de estudantes obterá o mesmo resultado. Já aqueles que esperam uma variabilidade natural ou acidental nos resultados experimentais tenderiam a responder que o próximo valor seria apenas parecido com o primeiro valor obtido.

A questão 3 expressa que, ao repetir a coleta de dados, com o mesmo valor de 20 cm para a altura h , o mesmo grupo de estudantes obteve um valor de 66,8 cm. É solicitado então que o aluno comente sobre o resultado obtido. Assim, deseja-se saber se o respondente encara o segundo resultado com naturalidade, conseguindo atribuir alguma causa ou motivo para a variação apresentada. Outra reação possível é negar o valor obtido, argumentando que houve algum equívoco ou erro de procedimento, pois os valores deveria ser os mesmos.

Os dados para este trabalho foram coletados no início do ano, para se identificarem os conhecimentos iniciais, e ao final do ano letivo, para que fosse possível identificar modificações na compreensão dos participantes, constituindo dessa forma um pré-teste e um pós-teste. A coleta do material empírico da pesquisa ocorreu durante o horário das aulas de física e ficou sob a responsabilidade dos próprios professores de física das turmas, os quais possuem experiência em pesquisas educacionais. Os alunos responderam aos testes individualmente.

Além da realização dos testes, foram registradas, em áudio e vídeo, as atividades de alguns grupos de alunos no laboratório durante 7 das 13 atividades (sendo a 3ª, 4ª e as 5 últimas atividades). Também foram realizadas entrevistas individuais com 9 estudantes e com 7 grupos de alunos (grupos de 2 a 6 pessoas, selecionados de acordo com a disponibilidade e voluntariedade). Os alunos puderam decidir se preferiam ser entrevistados individual ou coletivamente. As entrevistas foram realizadas após a 8ª atividade e ao final do ano letivo, durante as quais foram feitas perguntas sobre as questões respondidas nos testes e sobre aspectos do laboratório relacionados à pesquisa. Em três delas, foi realizada em grupo uma atividade de lançamento de projétil com os alunos para contextualizar as questões discutidas. Por meio das entrevistas, buscou-se esclarecer melhor as concepções dos participantes sobre a reprodutibilidade dos dados experimentais, discutir as respostas apresentadas nos testes e obter suporte empírico para as inferências durante a análise dos dados.

As questões abaixo se referem a um experimento ilustrado na figura ao lado, realizado por grupos de estudantes durante uma aula no laboratório de Física. Uma rampa de madeira é colocada na extremidade de uma mesa. Pode-se soltar uma pequena esfera de uma altura h (com relação a mesa). A esfera percorre toda a rampa e alcança o chão a uma distância d da borda da mesa.

Um papel especial é colocado no chão, sobre o qual a esfera, ao cair, deixa uma pequena marca para que os estudantes possam localizar onde a esfera caiu. Os estudantes estão investigando como a distância d varia em função da altura h . Uma fita métrica é utilizada para as medições.

1) Um grupo de três alunos discute sobre as estratégias para coletarem os dados para alcançar o objetivo proposto. Eles decidem que coletarão os dados para cinco alturas h diferentes (10, 15, 20, 25, 30 cm). Porém, discordam sobre as medições da distância d para cada altura. Abaixo, estão as opiniões dos alunos. Qual a opinião que você considera mais adequada? **Justifique sua resposta.**

Estudante 1 – Soltaremos a esfera de cada uma das alturas h e mediremos a distância alcançada por ela apenas uma vez. Uma medida da distância para cada altura basta.
 Estudante 2 – Faremos duas medidas da distância alcançada pela esfera para cada altura.
 Estudante 3 – Faremos pelo menos 5 medidas da distância alcançada pela esfera para cada altura.

2) Um outro grupo de alunos começa a atividade e solta a esfera de uma altura h de 20 cm. Eles medem a distância d alcançada pela esfera ao atingir o chão e obtêm o valor de 67,4 cm. Eles resolvem realizar nova coleta de dados com a mesma altura. Qual o valor que você espera que eles obtenham? **Justifique sua resposta.**

3) Ao repetir a coleta de dados, com o mesmo valor de 20 cm para a altura h , o mesmo grupo de estudantes obteve um valor de 66,8 cm. Comente sobre o resultado obtido pelo grupo de estudantes.

Figura 1 – Instrumento de pesquisa utilizado para a coleta de dados.

RESULTADOS E DISCUSSÃO

Para identificar as concepções dos alunos sobre o tamanho da amostra e se eles reconheciam a necessidade de se realizarem medidas repetidas, analisaram-se suas respostas à questão 1 do questionário aplicado. A Tabela 1 apresenta a distribuição das escolhas dos participantes no pré e pós-teste de acordo com três comentários.

Tabela 1 – Distribuição das respostas dos participantes para a primeira questão do instrumento de pesquisa, de acordo com os comentários disponíveis.

Comentário	Número de medidas	Pré-teste		Pós-teste	
		N	%	N	%
Estudante 1	Apenas 1	21	12,8	3	1,8
Estudante 2	2 medidas	18	11,0	8	4,9
Estudante 3	Pelo menos 5	125	76,2	153	93,3
Totais		164	100	164	100

A maioria dos alunos, já no pré-teste, percebeu a necessidade de se realizarem mais medidas. Dos 164 alunos, 125 (76,2%) optaram corretamente pelo comentário do estudante 3. Essa consciência foi acentuada durante a realização das atividades ao longo do ano. No pós-teste, a porcentagem de alunos que

concordaram com o estudante 3 aumentou para 93,3%. Contrastaram-se, individualmente, as respostas dos alunos nos dois testes. Dessa forma, 117 alunos concordaram com o estudante 3 em ambos os testes. Dos 47 alunos restantes, 36 (76,6%) mudaram de opinião e concordaram com o estudante 3 no pós-teste, o que resultou numa diferença significativa entre os testes – MH (164) = 4,308, $p < 0,001^1$.

Analisaram-se também as justificativas utilizadas para sustentar a escolha do comentário. Observando os padrões de respostas fornecidas pelos alunos e recorrendo a algumas categorias criadas por Lubben *et al.* (2001), desenvolveu-se uma categorização para as justificativas dos participantes. Para cada um dos três comentários, fez-se necessário desenvolver categorias específicas, semelhantes às propostas por Kok e Priemer (2023). As demais categorias se aplicam a mais de um comentário.

Justificativas para a escolha pelo comentário do estudante 1 – apenas uma medida

Em geral, os alunos que optaram pelo comentário do estudante 1 não reconhecem a necessidade de se realizar mais que uma medida, porém, as justificativas divergem. Enquanto alguns acreditam que o resultado será sempre o mesmo, outros defendem que os resultados podem divergir quando se coletam mais dados. Portanto, realizar apenas uma medida evitaria dúvidas quanto ao valor 'correto'.

– Não há necessidade de se repetirem medidas: deixa explícito que a coleta de apenas um dado para cada altura seria o suficiente, não havendo necessidade de se repetirem medidas. Além disso, justifica afirmando que se repetir o processo, obterá sempre o mesmo resultado. Exemplos:

“Porque se o experimento for feito de forma adequada, cada altura terá apenas uma distância possível e, todas as vezes que o experimento for repetido eles encontrarão a mesma medida.”

“Uma medida basta, pois, se soltar a bolinha de uma mesma altura várias vezes, sempre obterá o mesmo resultado de distância.”

– Evitar erros: nessa resposta, percebe-se que o aluno sabe que é normal a obtenção de resultados diferentes, mas ele afirma que o experimento deve ser feito com atenção e cuidado, coletando apenas um dado para evitar erros ou diferenças entre os valores. Exemplos:

“Pois cada vez que vamos medir a distância o resultado será um pouco diferente (por causa de erro na colocação da esfera etc.), então devemos fazer 1 vez bem feito.”

“Porque para observar se a altura varia em função da distância, basta medir uma vez, porque se for medido mais de uma vez outros fatores podem interferir nos resultados.”

Justificativa para a escolha pelo comentário do estudante 2 – duas medidas

Geralmente, os alunos que optaram pelo comentário do estudante 2 reconhecem a necessidade de se coletarem mais dados, mas acreditam que apenas duas medidas sejam suficientes.

– Dois medidas são suficientes: apesar de afirmarem que duas medidas seriam suficientes, os alunos não deixam claro o que fariam com os dois dados coletados. Na maioria das vezes, argumentam que uma medida é pouco, mas que 5 são desnecessárias. Exemplos:

“Porque não seriam necessárias mais de duas medidas.”

“Pois não deve ser realizada apenas 1 medida como o Estudante 1 disse, nem são necessárias tantas medidas quanto o Estudante 3 disse.”

Justificativas para a escolha pelo comentário do estudante 3 – cinco medidas

Os alunos que optaram pelo comentário do estudante 3 reconhecem que 5 medidas seria o mais indicado para a situação, argumentando, inclusive, que a média dos valores obtidos seria mais precisa e os

¹ Para verificar a significância das mudanças no entendimento dos participantes da pesquisa, utilizou-se o teste de homogeneidade marginal, que é similar ao teste de McNemar, porém, estendido aos casos nos quais a variável de interesse assume mais de dois valores nominais (Agresti, 2013). Para cada teste realizado, reportam-se o valor padronizado da estatística do teste e o nível de significância obtido.

resultados mais assertivos, diminuindo os erros. Porém, em grande parte das justificativas, não está clara a real importância de se repetir a medida várias vezes ou como esses números contribuiriam para a obtenção de resultados mais precisos ou para a diminuição dos erros.

– Obter uma média mais precisa: afirmam que com mais dados é possível a determinação de uma média mais precisa. Porém, não apresenta indícios sobre como ou por que o cálculo da média com mais dados seria mais ou menos preciso que o cálculo efetuado com uma menor quantidade de dados. Exemplos:

“Pois quanto maior o número de medidas, mais preciso será o valor da média dos valores medidos.”

“Quanto mais medidas você obter, maior precisão a média resultante terá.”

– Obter resultados mais seguros ou precisos: afirmam que com mais dados seria possível se obterem resultados mais precisos ou seguros, porém não apresentam justificativas do porquê ou de como esses dados seriam analisados. As justificativas desta categoria se distinguem daquelas categorizadas como “obter uma média mais precisa” por não se referirem à média. Exemplos:

“Quanto mais medidas melhor para obtermos uma precisão maior quanto ao resultado.”

“Pois quanto mais medidas da distância eles fizerem por altura, mas exatidão eles terão no resultado.”

– Diminuir os erros: nessa categoria, nota-se que os alunos atribuem a necessidade de se coletarem mais dados à diminuição dos erros, sem, no entanto, apresentarem um motivo para tal. Exemplos:

“Porque quando estamos fazendo algum experimento, temos que repetir a mesma coisa várias vezes, com isso nós diminuimos a possibilidade de erros muito graves.”

“Com mais tentativas diminui as chances de erro da experiência.”

Justificativas para a escolha pelos comentários dos estudantes 2 e 3

As justificativas abaixo estão presentes nas respostas dos alunos que optaram pelos comentários dos estudantes 2 e 3.

– Conferir o resultado: o aluno afirma que é necessária a repetição das medidas para a confirmação ou a conferência do resultado. Além disso, muitas vezes ele diz que os demais resultados serão iguais. Exemplos:

“Não era muitos fatores que podem alterar o resultado, mas é sempre bom conferir”.

“Pois assim irá conferir para ver se o resultado é o mesmo.”

– Calcular a média: afirma claramente que o objetivo de se medir mais vezes uma determinada variável seria calcular a média dos valores obtidos. Contudo, não há menção sobre a importância da média ou sobre alguma medida de dispersão dos dados. Exemplos:

“Com 5 resultados, pode-se tirar a média das medidas, obtendo uma resposta confiável.”

“Pois assim teríamos mais dados e poderíamos fazer um média da distância d , achando um resultado mais confiável em cada altura.”

Na Tabela 2, estão distribuídas as respostas dos alunos em função dos comentários e das categorias de cada comentário. No pré-teste, o aluno que optou pelo comentário do estudante 1 tendeu a justificar sua escolha argumentando que não há necessidade de se repetir. Pode-se fazer um paralelo entre essa categoria e os níveis A e B (ver Quadro 1) identificados por Lubben e Millar (1996), o que indica uma concepção muito restrita do processo de medição. Já no pós-teste, o número de alunos que optou por esse comentário é muito reduzido.

Os alunos que optaram pelo comentário do estudante 2, tanto no pré-teste quanto no pós-teste, tiveram a maioria das justificativas classificadas como “conferir o resultado” ou “duas medidas são suficientes”. Os alunos que optaram pelos comentários dos estudantes 1 e 2 revelam indícios de pensamento segundo o ‘paradigma pontual’, proposto por Allie e colaboradores (1998).

Os dados da Tabela 2, assim como os da Tabela 1, nos permitem verificar que no pós-teste, há uma diminuição significativa dos alunos que optaram pelos comentários dos estudantes 1 e 2, indicando uma relativa sofisticação nas concepções sobre o processo de medição. Entretanto, pela análise das justificativas dos alunos que optaram pelo comentário do estudante 3, tanto no pré-teste quanto no pós-teste, percebe-se que não houve o real entendimento sobre a necessidade de se repetirem as medidas. A ocorrência das categorias “calcular a média” e “obter média mais precisa” revela que os alunos aprenderam operacionalmente sobre a necessidade de se coletarem mais dados e efetuar o cálculo da média. Porém, em resposta alguma do conjunto de dados se percebe uma concepção abrangente sobre a importância do cálculo da média e a utilização de alguma medida de dispersão dos dados para se avaliarem os resultados. A premissa apresentada na categoria “obter média mais precisa” – com mais dados se pode obter uma média mais precisa – nem sempre é verdadeira. Se os dados adicionais não forem coletados com os mesmos cuidados e a mesma qualidade dos demais, pouco acrescentarão. No entanto, o comentário procede com base na suposição de que não há erros sistemáticos, apenas erros aleatórios. No laboratório, os alunos raramente praticam atividades envolvendo erros sistemáticos ou erros deliberados de procedimento. Assim, há uma suposição implícita de que apenas os erros aleatórios estão presentes nas atividades que realizam.

Tabela 2 – Distribuição das justificativas dos participantes para a resposta à primeira questão do instrumento de pesquisa, de acordo com a categorização elaborada.

Comentário	Justificativas	Pré-teste		Pós-teste	
		N	%	N	%
Estudante 1	Não há necessidade de repetir	16	9,8	1	0,6
	Evitar erros	2	1,2	1	0,6
	Outras	3	1,8	0	0,0
	Nenhuma	0	0	1	0,6
	Subtotal	21	12,8	3	1,8
Estudante 2	Duas medidas são suficientes	5	3,1	5	3,1
	Conferir o resultado	8	4,9	3	1,8
	Outras	5	3,0	0	0
	Subtotal	18	11,0	8	4,9
Estudante 3	Calcular a média	45	27,4	45	27,4
	Obter média mais precisa	8	4,9	22	13,4
	Obter resultados mais precisos	35	21,3	27	16,5
	Diminuir os erros	10	6,1	29	17,7
	Conferir o resultado	6	3,7	0	0
	Outras	21	12,8	30	18,3
	Subtotal	125	76,2	153	93,3
Total	164	100	164	100	

Durante a maioria das atividades práticas realizadas ao longo do ano, os alunos tiveram de calcular a média de alguns valores. Além disso, o número de 5 medidas parece ser um “número mágico” no laboratório, ou um número padrão, pois é utilizado em muitas atividades. Por isso, parece fácil para os alunos aprenderem que seria necessária essa quantidade de dados, aprendendo assim, por repetição, que após coletarem os dados, devem calcular a média. Isso justifica a importância exagerada que os alunos atribuem à média. Porém, eles não compreendem bem nem o que a média dos valores obtidos realmente significa, nem o que ela representa. Exemplos:

“Fazer várias vezes o experimento e tirando a média dos valores elimina erros de medida.”

“Os estudantes devem coletar mais dados e tirar a média. Assim os erros são eliminados.”

Nas entrevistas, notou-se que grande parte dos alunos sabe que deve repetir as medições um certo número de vezes, o que, sob certas condições, leva conseqüentemente a um resultado melhor, porém, desconhecem as razões disso. Os três trechos a seguir foram extraídos das entrevistas realizadas com grupos de alunos em uma atividade que envolvia lançamento de projéteis, abordando o assunto:

P- “O aluno A2 sugeriu que fizéssemos 5 medidas. E se alguém se propusesse a fazer 15 medidas?”

A2- “A precisão ia ser maior.”

A6- “Um resultado mais confiável.”

P- “São sinônimos? Resultado mais preciso e confiável?”

A6- “Eu não sei explicar, mas acho que são.”

A1- “Quando se tem uma coisa mais precisa, conseqüentemente é uma coisa mais confiável. Se é uma coisa mais precisa, é uma coisa mais próxima do que aconteceu.”

A2- “Com mais dados, o resultado vai ser muito mais próximo do real.”

A3- “Porque é mais preciso. A média vai estar mais de acordo com a realidade.”

Os alunos acreditam que um maior número de medidas levaria a resultados mais precisos e confiáveis, mas não sabem articular as razões para isso.

P- “Se eu perguntasse para vocês, qual a distância percorrida pela esfera? Como vocês me responderiam?”

A3- “Tirava uma média desses dados.”

A2- “Faria outras medidas para garantir.”

P- “Garantir o que?”

A3- “Uma precisão maior.”

A2- “Para a média ficar mais detalhada.”

P- “Vocês acham que quanto maior o número de dados melhor?”

A1- “Fica mais próximo do real.”

P- “Quantos dados vocês coletariam?”

A1- “10 medidas.”

P- “Vocês então fariam 10 medidas e depois, o que fariam com essas medidas?”

A2- “Tirava a média.”

P- “O que representa a média?”

A2- “Um valor em comum entre todas.”

A1- “Um valor que vai se aproximar de todas.”

P- “Por que vocês calculam a média?”

A3- *“Porque se eu pegar um valor será aleatório. Se eu pegar a média, estou pegando um valor comum a todos.”*

A2- *“Porque a gente não considera o erro. Se a gente fizer só uma tentativa, você parte do pressuposto que essa tentativa está certa. Se fizer várias e receber resultados diferentes, você começa a considerar a possibilidade de erros.”*

P- *“Mas então seria melhor fazer apenas uma medida. Assim o experimento não teria erro.”*

A1- *“Erro vai ter.”*

A2- *“Não é que fazendo apenas uma vez não vai ter erro, você não vai saber qual o erro.”*

A3- *“Não vai ter um valor para comparar.”*

Pode-se observar que A1 e A2 não conseguem articular uma resposta satisfatória sobre o que representaria a média de um conjunto de dados. Os três reconhecem a possibilidade de erros durante a experimentação. A2 e A3 apresentam indícios que compreendem que, dado um conjunto de medidas, seria possível efetuar comparações para estabelecer a qualidade dos dados e calcular um parâmetro indicador do “erro” existente.

P- *“Se eu perguntasse para vocês, qual a distância percorrida pela esfera? Como vocês me responderiam?”*

A1- *“Faria muitos lançamentos e tiraria a média.”*

P- *“Muitos lançamentos? Quantos?”*

A2- *“Cinco.”*

A3- *“Quanto mais, melhor.”*

A4- *“Duzentos e cinquenta.”*

P- *“Porque quanto mais lançamentos é melhor?”*

A3- *“Porque a média fica mais exata. Dá para ver os valores máximos e mínimos.”*

A4- *“Porque vai ter a repetição de um monte de resultado.”*

P- *“É bom ter a repetição dos resultados?”*

A4- *“É bom. Porque o seu resultado final vai se aproximar disso.”*

P- *“Mas o que é a média de um conjunto de medidas?”*

A5- *“É a soma das medidas, dividida pelo número de medidas”.*

P- *“Essa é a definição matemática. Mas o que representa a média?”*

A2- *“O valor mais preciso da medida. O valor que mais representa a medida.”*

P- *“Qual a diferença entre a média com cinco medidas e com duzentas e cinquenta?”*

A3- *“A média com duzentas e cinquenta é mais precisa.”*

P- *“E em relação ao erro?”*

A6- *“Quanto mais medidas, bem maior o erro, pois cada medida tem um errinho.”*

Esse trecho mostra, novamente, que os alunos percebem que é melhor coletar o maior número de medidas possíveis, mas não conseguem articular as razões para isso. Muitos alunos (como o A4) defendem

que seria para se obterem vários valores repetidos e outros (como o aluno A3) acreditam que a média ficaria mais exata ou precisa.

Quando começam a realizar atividades práticas, os alunos entram em contato com diversas fontes de erro. Porém, eles não nem conseguem distingui-las nem as denominar corretamente, o que resulta em compreensões incorretas, como a da aluna A6, no trecho acima. As respostas categorizadas como “diminuir os erros” indicam justamente que os alunos consideram erros de medida, erros de procedimentos e interferência de fatores externos, entre outros, como ‘erros’. Tal concepção prejudica a compreensão correta do significado da média, já que seu cálculo elimina apenas os erros aleatórios de medida, mas não elimina a influência de outros fatores, erros sistemáticos ou falhas na execução da atividade.

A elevada percentagem da categoria “obter resultados mais precisos” também pode indicar algo semelhante. Como pode ser visto nos exemplos, as respostas para essa categoria são vagas e podem ser interpretadas de diversas maneiras. O aumento da precisão e a obtenção de resultados melhores são atribuídos a um posterior cálculo para a média (categoria “obter média mais precisa”) ou os dados subsequentes são melhores que os anteriores, uma vez que os alunos estariam adquirindo prática (semelhante ao nível C de Lubben e Millar (1996)). Veja os exemplos abaixo:

“Em um experimento o melhor é coletar o maior número de dados possíveis para que se chegue a uma medida próxima da exatidão.”

“Quanto mais vezes você repete o experimento, mais você aproxima do resultado certo.”

Devido à ampla ambiguidade, essas respostas não foram categorizadas como “obter média mais precisa” e podem indicar concepções incorretas sobre o processo de medição. Os alunos parecem acreditar que existe um valor ‘correto’ para a grandeza medida e que ele poderia ser obtido por meio do cálculo da média ou tomando-se o valor mais frequente. Identificamos também respostas coerentes com o nível D, ou seja, devem-se coletar mais dados e escolher aquele que mais se repetiu. Exemplos:

“Eles devem fazer as 5 medidas e pegar aquela que mais aparece.”

“É bom eles repetirem mais vezes para ver se a medida vai se repetir.”

Dessa forma, a escolha pelo comentário do estudante 3 não implica, necessariamente, que o aluno tenha compreendido a necessidade e a razão de se coletarem mais dados durante um experimento, bem como o que se fazer com os dados coletados. As respostas dos alunos nas entrevistas exemplificam isso:

P- “O que significa a média? Qual a sua importância?”

A1- “Para você tirar a chance de erro.”

A3- “Porque existem os erros de medida e se calcula a média para ficar mais preciso.”

A2- “Porque toda medida tem um erro de medida. Fazendo mais medidas tira-se a média, que vai ser o valor mais representativo.”

P- “Por que precisa medir mais vezes?”

A3- “Você diminui os erros de medida.”

A1- “É muito difícil ser 100% igual. As medidas variam.”

A2- “Com mais medidas, a média é mais exata. A precisão vai ser maior.”

P- “Da onde vem essa precisão?”

A1- “A precisão é melhor porque se mede mais vezes.”

A2- “Não sei te explicar.”

Também nesse trecho, vê-se que os alunos reconhecem a necessidade de se coletarem várias medidas e calcular a média, mas, novamente, desconhecem o real significado da média e o porquê de se coletarem várias medidas. Da mesma forma, o trabalho de Allie e colaboradores (1998) demonstrou que muitos alunos sugeriram que vários dados deveriam ser coletados, porém, não compreendem as razões para isso. Muitos afirmaram, assim como nesta pesquisa, que a razão é se obter ou calcular a média, sem explicitar, no entanto, qual seria o benefício desse cálculo para a análise dos dados. O diálogo a seguir, gravado durante a realização da última atividade no laboratório, ilustra bem esse cenário:

P- “Por que nós repetimos as medidas quando realizamos as práticas?”

A1- “Para eliminar os erros. As medidas que fazemos tem diferenças.”

P- “Por que existe essa diferença?”

A1- “Porque nunca vai ser exato.”

P- “Por quê?”

A1- “Sei lá!”

A2- “É mesmo! Por que dá essa diferença? A gente sabe que dá essa diferença por isso a gente precisa de mais dados.”

P- “Mas nem com um instrumento muito preciso?”

A2- “Mesmo se o instrumento for bom, sempre haverá erros de medida.”

P- “E por que precisa de mais medidas?”

A2- “Quanto mais vezes você fizer mais chance você terá de fazer a média.”

P- “O que representa a média? O que ela pode te indicar?”

A2- “A média das medidas. Quando você tira a média, você quer saber mais ou menos uma proporção básica, o que tem em comum a todos.”

A1- “Ah... é muito difícil explicar isso, por isso não vou fazer Física.”

Os resultados de Rollnick, Dlamini, Lotz e Lubben (2001) corroboram os resultados apresentados até aqui. Também por meio de um questionário, os autores investigaram as concepções de universitários sul-africanos sobre as razões de se repetirem as medidas realizadas durante um experimento. A maioria dos participantes se dividiu em três grupos, cujas razões para se repetirem os experimentos são: a obtenção de valores recorrentes (nível D); a prática (nível C) e o cálculo da média (nível F). Além desses grupos, 9% responderam que não seria necessária a repetição das medidas (níveis A e B). Apesar de 18% responderem que a razão da coleta de mais dados seria para possibilitar o cálculo da média, o desempenho desses alunos em outras questões revela que eles consideram o cálculo da média um mero exercício matemático, não atribuindo a tal média uma medida de tendência central ou a relacionando-a à dispersão dos dados, conforme identificado por Séré *et al.* (1993).

Pesquisaram-se também as concepções dos alunos sobre a reprodutibilidade dos dados durante a realização de uma atividade experimental. Para isso, analisaram-se as respostas fornecidas para a questão 2, as quais puderam ser classificadas, basicamente, em duas categorias:

– Mesmo valor: o aluno diz que o resultado obtido será rigorosamente igual ao anterior. Para esses alunos, a obtenção do mesmo valor indica que o experimento foi realizado corretamente e, pela mesma lógica, a obtenção de valores distintos é indicativo de erro nos procedimentos. Exemplos:

“Eu espero que a distância não se altere porque a física é exata. Se mudar é porque alguma coisa deve ter sido feita errada.”

“O mesmo valor deve ser obtido, pois assim saberemos se eles estão fazendo corretamente o procedimento.”

– Valor próximo: o aluno afirma que o resultado será próximo ao primeiro valor obtido. Exemplos:

“Eu espero que eles obtenham um valor diferente de (67,4 cm), porque sempre ocorre um erro de medida.”

“Um valor próximo a 67,4 cm, por ser o valor encontrado em sua primeira marcação e porque normalmente não se consegue um mesmo valor.”

Essa resposta foi fornecida por 158 alunos nos dois testes² e os resultados constam na Tabela 3. No pré-teste, 122 (77,2%) alunos afirmaram que o grupo obterá um valor próximo ao anterior e outros 23 (14,3%) que o valor seria idêntico ao que foi previamente obtido. No pós-teste, 144 (91,1%) respondentes afirmaram que o valor seria próximo ao primeiro, enquanto 9 (5,7%) afirmaram que o grupo obterá um valor idêntico ao primeiro. Ao se contrastarem individualmente as respostas dos alunos nos dois testes, verifica-se que não houve diferenças significativas – MH (158) = 0,866, $p = 0,472$. Dos 158 alunos, 115 responderam corretamente em ambos. Dos 43 restantes, 29 (67,4%) responderam corretamente no pós-teste.

Tabela 3 – Distribuição das respostas dos participantes para a segunda questão do instrumento de pesquisa, de acordo com as categorias estabelecidas.

		Pós-teste			
Respostas		Valor próximo	Mesmo valor	Outras	Total
Pré-teste	Valor próximo	115	3	4	122
	Mesmo valor	20	3	0	23
	Outras	9	3	1	13
Total		144	9	5	158

Procurou-se identificar se haveria uma relação entre a escolha pelo comentário na questão 1 e a resposta para a questão 2. Nesse caso, consideraram-se apenas os resultados do pós-teste. A Tabela 4 apresenta essa relação.

Tabela 4 – Evidência da relação entre as respostas dos participantes no pós-teste para as duas primeiras questões do instrumento de pesquisa.

Categorias		Questão 1			Totais
		Estudante 1	Estudante 2	Estudante 3	
Questão 2	Mesmo valor	1	0	8	9
	Próximo	2	8	134	144
Totais		3	8	142	153^(*)

(*) O total de 153 respondentes foi obtido excluindo-se 5 respondentes (do total de 158) que tiveram suas respostas categorizadas como ‘outras’.

Percebe-se que os alunos que optaram pelo comentário do estudante 3 na questão 1 tenderam a responder que o próximo dado coletado seria parecido, mas não igual ao anterior. Contudo, algumas células da tabela apresentam incoerências conceituais. Se o aluno concorda com o comentário do estudante 1 (apenas uma medição), era de se esperar que acreditasse que o próximo resultado seria idêntico ao primeiro. Ao mesmo tempo, o aluno que concorda com o comentário do estudante 3 (5 medições), não poderia esperar que o próximo resultado fosse idêntico ao primeiro. Assim, alguns dados da Tabela 4 reforçam as inferências sobre as concepções fragmentadas dos alunos sobre a reprodutibilidade dos dados e a necessidade de repetição das medições em uma atividade experimental. Portanto, além de demonstrar a forte relação entre os dois itens analisados, os dados da tabela também contribuem para demonstrar que

² A diferença entre o número de respondentes da primeira questão para a segunda se deu pelo fato de que alguns alunos deixaram essa resposta em branco, tanto no pré como no pós-teste.

o conhecimento de alguns alunos sobre o processo de medição, ao final do ano letivo, ainda é fragmentado e não integrado num sistema conceitual coeso, conforme proposto por Camargo Filho *et al.* (2015).

Na questão 3, afirmou-se que ao repetir a coleta de dados, o mesmo grupo de estudantes obteve um valor de 66,8 cm. Em seguida, solicitou-se aos alunos comentassem sobre o resultado obtido pelo grupo. A análise das respostas permitiu classificá-las em 6 categorias:

– Resultado normal: o resultado é considerado dentro do esperado, já que não se diferenciou muito do anterior. Em muitos casos, reconhece-se a ocorrência de erros, justificando a variação do valor. Exemplos:

“É um número próximo de 67 cm, porém variações podem ocorrer, o atrito da esfera pode ter sido maior, existem muitas hipóteses para que o resultado encontrado seja este.”

“Pelo menos para mim, o resultado era esperado, pois qualquer procedimento realizado novamente está sujeito a variações, pois sempre ocorrem modificações, nunca são idênticos.”

– Erros experimentais ou de medida: afirma-se que algum tipo de erro aconteceu, seja ele de medida, experimental ou procedimental. Porém, não fica claro se o resultado foi considerado normal ou não. Exemplos:

“Algo pode ter dado errado, talvez a altura tenha sido medida errado ou os alunos não souberam usar a fita métrica direito.”

“Acho que eles soltaram a bola do valor de 15 cm, por isso a bola percorreu menos centímetros que a outra.”

– Resultado anômalo: responde explicitamente que o resultado obtido deveria ser idêntico ao anterior. Afirma que o experimento foi realizado em condições idênticas e que não há erros. Exemplos:

“Na minha opinião fiquei um pouco confusa pois se a altura é a mesma, porque deu um valor diferente?”

“Por soltarem na mesma altura, o resultado deveria ser o mesmo, já que não variou a altura.”

– Variações devido a fatores externos: atribuem as causas da variação a fatores externos como resistência do ar, atrito e outros. Também não deixam claro se consideram a variação normal, mas pode-se dizer que não atribuem a diferença a erros de medida. Exemplos:

“Provavelmente algum fator como atrito atrapalhou na nova experiência.”

“Pode ter havido algum impulso ou influência do meio como vento, respiração, pressão etc.”

– Resultado diferente: consideram o segundo resultado bem diferente do primeiro. Há diferença entre os resultados e, implícita ou explicitamente, indicam que houve algum erro de procedimento. Exemplos:

“Os estudantes se enganaram ao repetir os experimentos, pois os resultados foram bem distintos.”

“É possível concluir que uma das medidas está incorreta, pois a margem de uma para outra foi muito grande.”

– Dados insuficientes: afirma que apenas dois dados não seriam suficientes para afirmar sobre a similaridade/disparidade entre os resultados, sendo, portanto, necessário se obterem mais valores. Exemplos:

“Não sabemos se o primeiro ou o segundo está mais certo, seriam necessárias mais marcações.”

“Esse resultado mostra que há uma ‘margem de erro’ e que eles devem fazer mais experimentos”.

Inicialmente, é interessante se discutir a ocorrência da categoria “resultado diferente”, na qual os alunos afirmam que o resultado obtido é significativamente diferente do anterior. A diferença entre os valores é de apenas 0,6 cm, ou aproximadamente 1%. Alguns afirmaram que só o algarismo avaliado poderia variar, sendo que se algum outro variasse, o resultado seria diferente.

A Tabela 5 apresenta a distribuição das respostas dos 158 alunos que responderam à questão nos dois testes, de acordo com a categorização criada. Verifica-se que não há diferenças significativas entre o desempenho dos alunos no pré-teste e no pós-teste. Se levarmos em consideração que apenas as respostas categorizadas como “resultado normal” e “dados insuficientes” deixam claro que não havia grandes diferenças nos valores, ou que não se poderia afirmar nada, juntas somando 53,8% no pós-teste, grande parte dos alunos, mesmo tendo respondido no item anterior que o resultado esperado deveria ser próximo do primeiro, teve problemas para avaliar e justificar o novo resultado. De qualquer forma, a presença da categoria “dados insuficientes” foi uma surpresa positiva. Tais respostas demonstraram certa sofisticação no raciocínio ao responderem, adequadamente, que não poderiam afirmar nada sobre a questão, já que seria necessário coletar mais dados para poderem avaliar os resultados.

Tabela 5 – Distribuição das respostas dos participantes para a questão 3 do instrumento de pesquisa, de acordo com as categorias estabelecidas.

Categorias	Pré-teste		Pós-teste	
	N	%	N	%
Resultado normal	75	47,5	81	51,3
Erros experimentais ou de medida	43	27,2	49	31,0
Resultado anômalo	7	4,4	0	0,0
Variações devido à fatores externos	9	5,7	6	3,8
Resultado diferente	8	5,1	12	7,6
Dados insuficientes	2	1,3	4	2,5
Outras	14	8,9	6	3,8
Totais	158	100	158	100

Outro aspecto que se pode extrair da Tabela 5 é que os alunos não apresentam dificuldades de reconhecer que existem algumas fontes de erros durante o processo de medição que influenciam o resultado (categorias “erros experimentais ou de medida” e “variações devido a fatores externos”). Mesmo as justificativas daqueles que consideraram o resultado normal continham alusões às fontes de possíveis variações para os resultados. Porém, mesmo reconhecendo a existência de erros ou fatores que poderiam influenciar nos resultados, os alunos demonstraram dificuldades de avaliarem se a variação obtida poderia ou não ser fruto desses erros. Além disso, como foi abordado no item anterior, percebe-se certa confusão dos alunos ao se referirem aos erros, demonstrando que não possuem ainda um bom entendimento sobre os possíveis tipos de erros que poderiam ocorrer num experimento.

Ao começarem a ter contato com as práticas de laboratório, é normal que os alunos atribuam quaisquer variações nos resultados a erros. Porém, muitos (conforme se percebe nas respostas) não compreendem bem, por exemplo, a diferença entre erros sistemáticos e aleatórios ou erros devido a fatores externos etc. Outra dificuldade dos alunos é quantificar a influência desses fatores e dos erros presentes no experimento.

A análise das respostas nos testes e nas entrevistas revela que poucos alunos entendem bem os pressupostos em relação aos tipos de erros e à irreprodutibilidade dos fenômenos físicos (de que é impossível se repetir um experimento exatamente nas mesmas condições do anterior e obter exatamente o mesmo resultado, por exemplo). Sendo assim, as variações nos valores medidos não devem ser atribuídas apenas a erros de medida ou na execução.

Durante as entrevistas, foi perguntado aos alunos por qual razão os valores medidos num experimento raramente coincidem. Destacam-se, aqui, apenas algumas respostas:

“Por que existem os erros acidentais de medida, os quais nos impedem de encontrar exatamente o mesmo valor.”

“Isso ocorre porque nenhum instrumento de medida é preciso, e podemos nos confundir ao olhar o objeto.”

“Porque vai depender muito do tempo de reação de cada pessoa que é diferente e também varia devido aos erros de medida que são comuns”.

“Porque constantemente ocorrem erros, sejam eles de medida ou ao realizar o procedimento.”

As repostas acima enfatizam que a baixa probabilidade de se obterem valores iguais para as medidas é atribuída apenas a erros de medidas ou erros de execução. De certa forma, alguns alunos revelaram possuir certa concepção sobre a irreprodutibilidade dos fenômenos físicos, justificando de outra maneira a variação das medidas obtidas durante a atividade experimental:

“Como na maioria dos experimentos também atuam outros fatores, muito dificilmente se obterá os mesmos resultados.”

“Porque existem uma série de fatores que atuam sobre um experimento, como é o caso do atrito, do vento etc.”

“Porque todo experimento envolve uma série de fatores que não dependem de quem está observando o experimento. Com a variação desses fatores, são obtidos resultados diferentes.”

Nas entrevistas em grupo, contextualizadas por meio da atividade de lançamento de projétil, alguns alunos não tiveram dificuldades de elencar algumas condições que interfeririam nos resultados. Abaixo, estão transcritos dois trechos, com dois grupos diferentes, que abordam as razões para a variação nos resultados.

P- “Se eu lançar a esfera novamente, o que vocês acham que vai acontecer?”

A1- “Vai cair próximo.”

A2- “Vai cair próximo, mas não vai cair no mesmo lugar.”

P- “Por quê?”

A2- “Erros no equipamento.”

A1- “Resistência do ar, o vento pode estar diferente.

A3- “A variação vai ser pequena.”

A2- “O que mais vai mudar vai ser o aparelho mesmo.”

P- “Não era para a esfera ter caído no mesmo lugar?”

A1- “Teoricamente sim...”

A2- “Uai, foram as mesmas condições, mesmo ângulo, tudo igual.”

P- “E por que a esfera não caiu no mesmo lugar?”

A2- *“Por causa da resistência do ar.”*

A3- *“A energia da mola da primeira vez pode ter sido diferente da segunda.”*

A2- *“Por causa de variações na mola.”*

Nas duas passagens fica claro que os alunos percebem que, em teoria, ou seja, se todas as condições permanecessem idênticas, a esfera cairia no mesmo lugar. Porém, como isso dificilmente ocorre, a esfera cairá em um lugar próximo, mas diferente do primeiro. Além disso, eles conseguem atribuir as variações obtidas a fatores plausíveis, todos relacionados à atividade. O domínio da concepção adequada sobre a reprodutibilidade de dados experimentais é importante para que o aluno possa entender e avaliar as variações que podem ocorrer nos resultados de um experimento para que possa, por exemplo, distinguir se a variação seria causada por qualquer tipo de erro em alguma fase da atividade ou se seria devido às relações causais entre as grandezas físicas envolvidas.

CONSIDERAÇÕES FINAIS

Os alunos demonstraram que aprenderam sobre a necessidade de se realizarem diversas medidas durante uma coleta de dados. A percentagem de alunos que acreditava que apenas uma ou duas medidas seriam necessárias diminuiu de 23,8% para 6,7%, e, conseqüentemente, a percentagem daqueles que reconhecem a necessidade de se realizar um número maior de medições chegou a 93,3% no pós-teste. Porém, ao se analisar suas justificativas, percebe-se que grande parte dos alunos não compreendeu ainda os motivos reais da importância de se trabalhar com um número maior de dados. Muitos argumentam que um maior número de medidas contribui para diminuir os erros ou para se obter dados mais seguros e precisos, mas não indicam como ou por que isso ocorre.

A inferência acima pode ser reforçada pelos resultados obtidos subsequentemente. Quando perguntados sobre qual deveria ser o próximo valor obtido com a repetição de uma medida (sendo que o valor da primeira medida era conhecido), novamente a grande maioria (91,1% no pós-teste) reconheceu que seria um valor próximo do primeiro, e não idêntico. Porém, quando solicitados a comentarem sobre o novo valor obtido (diferente de cerca de 1% do primeiro), apenas 53,8% deixaram explícito que o resultado seria normal. Boa parte dos participantes justificou a variação do resultado se referindo a erros experimentais ou de medida, sem deixar claro que a variação seria normal ou aceitável. Sendo assim, acredita-se que com a realização das práticas de laboratório, os alunos aprendem facilmente que devem coletar mais dados. Porém, a importância de se repetirem as medidas deve ser constantemente debatida e exemplificada para que eles realmente possam compreender suas ações no laboratório.

Outras pesquisas também avaliam as aprendizagens de conceitos de evidência relacionados ao processo de medição científica (Rollnick, Lubben, Lotz, & Dlamini, 2002; Kung & Linder, 2006). Essas pesquisas, apesar de reportarem melhoria no desempenho dos alunos no pós-teste, revelam que as aprendizagens alcançadas foram aquém do esperado. Para Pigozzo e Heidemann (2023, p. 332),

“sem instrução adequada e explicitamente dirigida para aspectos epistemológicos e procedimentais do processo de medição, os estudantes evoluem timidamente em suas concepções sobre a natureza da Ciência e desenvolvem poucas habilidades de coleta, processamento e análise de dados”.

A aprendizagem dos conceitos de evidência aquém do esperado pode ser explicada em parte pela distinção introduzida por Hart, Mulhall, Berry, Loughran e Gunstone (2000) entre os propósitos e os objetivos de uma atividade prática. Segundo eles, o primeiro diz respeito aos propósitos pedagógicos estabelecidos pelo professor, ou seja, a razão pela qual determinada atividade está sendo realizada, a forma como ela está organizada e relacionada a outras atividades e quais os resultados educacionais a atividade deveria produzir em termos de aprendizagem. Já os objetivos se referem aos enunciados mencionados para a turma, presentes nas folhas de atividades, ou seja, os objetivos são normalmente afirmações sobre um tópico específico de conteúdo, como, por exemplo, obter a relação entre tensão e corrente em condutores ôhmicos ou a relação entre a distensão e a força realizada em uma mola. Enquanto os objetivos da atividade podem ser vistos como algo que determina o que será feito durante a atividade, os propósitos se relacionam às expectativas dos professores de que a atividade resulte num aprendizado estabelecido, mas eles permanecem implícitos, às vezes até para os próprios professores. O grande problema é que os alunos reconhecem, na maioria das vezes, apenas os objetivos imediatos da atividade.

Portanto, os propósitos educacionais das atividades não são reconhecidos, o que os leva a não estabelecer relações nem entre as atividades realizadas nem entre as aprendizagens por elas geradas. Como foi destacado, os alunos tiveram três aulas no início do ano letivo e uma atividade prática na qual os conceitos de evidência relativos ao processo de medição foram trabalhados de forma explícita, como objetivo principal. No restante do ano, ao longo da maioria das atividades, esses conceitos foram abordados de forma implícita. Assim, muitos alunos tiveram dificuldades de reconhecer que a realização das práticas, além do aprendizado dos conteúdos específicos (constante elástica, relação tensão-corrente etc.), envolve também o aprendizado desses conceitos de evidência e das habilidades relacionadas à realização de atividades práticas no laboratório. Dessa forma, o desenvolvimento de instrumentos de coleta de dados como o utilizado neste trabalho é importante não apenas para a pesquisa em educação em ciências, mas para instrumentalizar o professor de ciências naturais para que possa abordar e avaliar os conhecimentos e habilidades de seus alunos em sala de aula.

Acredita-se, assim como Millar *et al.* (1994), que o processo de medição contém ideias importantes que devem ser ensinadas. Não se pode simplesmente esperar que os alunos cheguem a elas a partir das atividades. A compreensão de crianças e adolescentes sobre a evidência empírica e os critérios para avaliar a qualidade da evidência precisam ser trabalhados explicitamente pelo currículo. Raramente os alunos integram, espontaneamente, informações apresentadas isoladamente.

Ao longo do ano, os alunos realizaram atividades que proporcionavam condições de desenvolvimento de ideias, habilidades e formas de raciocínio relativas a diversos aspectos do processo de medição, como o cálculo do valor médio, avaliação da dispersão dos dados, identificação de possíveis fontes de erros de medida, representação dos dados em tabela, construção e análise de gráficos. Entretanto, apenas a realização de atividades práticas não é o suficiente para o desenvolvimento dessas ideias. Os professores precisam, durante as atividades, chamar a atenção dos alunos para os aspectos fundamentais do processo de medição, (tanto procedimentais quanto conceituais) para que construam uma compreensão adequada do processo de investigação científica. Ao integrarem os conceitos de evidência com concepções contemporâneas sobre a natureza e epistemologia da ciência, os alunos podem desenvolver uma perspectiva coesa sobre o processo de experimentação e sua importância para o desenvolvimento do conhecimento científico (Duschl, 2008; Osborne, 2016).

Mais pesquisas são necessárias para que se possa compreender melhor como ajudar os alunos a refletirem sobre o seu raciocínio e suas ações durante as atividades práticas, proporcionando-lhes meios de se perceber a importância da reflexão e da organização do seu pensamento, bem como suas ações para a resolução de problemas práticos. Deve-se investigar como as concepções dos alunos sobre o processo de medição evoluem ao longo de sua escolarização, acompanhando grupos por um período maior de tempo. Deve-se investigar, também, as concepções dos alunos sobre os diversos tipos de erros que podem ocorrer durante o processo de experimentação, em que fase do processo ocorrem, como podem ser evitados ou atenuados e como influenciam o resultado final. Dessa forma, será possível entender as dificuldades enfrentadas pelos alunos ao realizarem atividades práticas e desenvolverem atividades, propostas pedagógicas e metodologias de avaliação que contribuam para que estes tenham uma compreensão mais sofisticada sobre o processo de investigação.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro ao projeto APQ 00701-18.

REFERÊNCIAS

- Agresti, A. (2013). *Categorical data analysis*. (3rd Ed.). Hoboken, United States of America: John Wiley & Sons.
- Allie, S., Buffler, A., Campbell, B., & Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measurements. *International Journal of Science Education*, 20(4), 447-459. <https://doi.org/10.1080/0950069980200405>
- ANPEd - Associação Nacional de Pós-Graduação e Pesquisa em Educação. (2019). *Ética e pesquisa em Educação: subsídios*. Rio de Janeiro, RJ: ANPEd.

- Borges, A. T. (2002). Novos rumos para o laboratório escolar de ciências. *Caderno Brasileiro de Ensino de Física*, 19(3), 291-313. Disponível em <https://periodicos.ufsc.br/index.php/fisica/article/view/6607>
- Buffler, A., Allie, S., & Lubben, F. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*, 23(11), 1137-1156. <https://doi.org/10.1080/09500690110039567>
- Camargo Filho, P. S. D., Laburú, C. E., & Barros, M. A. D. (2015). Para além dos paradigmas da medição. *Ciência & Educação (Bauru)*, 21(4), 817-834. <https://doi.org/10.1590/1516-731320150040003>
- Coelho, S. M., & Séré, M. G. (1998). Pupils' reasoning and practice during hands-on activities in the measurement phase. *Research in Science & Technological Education*, 16(1), 79-96. <https://doi.org/10.1080/0263514980160107>
- Duschl, R. A. (2008). Science education in 3 part harmony: Balancing conceptual, epistemic and social goals. *Review of Research in Education*, 32, 268-291. <https://doi.org/10.3102/0091732X07309371>
- Gaya, T. F. M., & Bruel, A. L. (2019). Estudos longitudinais em educação no Brasil: revisão de literatura da abordagem metodológica e utilização de dados educacionais para pesquisas em Educação. *Revista de Estudos Teóricos y Epistemológicos en Política Educativa*, 4, 1-18. <https://doi.org/10.5212/retepe.v.4.015>
- Gott, R., & Duggan, S. (1995). *Investigative Work in the Science Curriculum*. Buckingham, United Kingdom: Open University Press.
- Gräber, W., Nentwig, P., Koballa, T. R., & Evans, R. H. (Eds.). (2013). *Scientific literacy: der Beitrag der Naturwissenschaften zur allgemeinen Bildung*. Berlin: Springer-Verlag.
- Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching*, 37(7), 655-675. [https://doi.org/10.1002/1098-2736\(200009\)37:7<655::AID-TEA3>3.0.CO;2-E](https://doi.org/10.1002/1098-2736(200009)37:7<655::AID-TEA3>3.0.CO;2-E)
- Kok, K., & Priemer, B. (2023). Assessment tool to understand how students justify their decisions in data comparison problems. *Physical Review Physics Education Research*, 19(2), 020141. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020141>
- Kranz, J., Baur, A., & Möller, A. (2023). Learners' challenges in understanding and performing experiments: a systematic review of the literature. *Studies in Science Education*, 59(2), 321-367. <https://doi.org/10.1080/03057267.2022.2138151>
- Kuhn, D., Arvidsson, T. S., Lesperance, R., & Corprew, R. (2017). Can engaging in science practices promote deep understanding of them? *Science Education*, 101(2), 232-250. <https://doi.org/10.1002/sce.21263>
- Kung, R. L., & Linder, C. (2006). University students' ideas about data processing and data comparison in a physics laboratory course. *Nordic studies in science education*, 2(2), 40-53. <https://doi.org/10.5617/nordina.423>
- Laburú, C. E., & Barros, M. A. (2009). Problemas com a compreensão de estudantes em medição: razões para a formação do paradigma pontual. *Investigações em Ensino de Ciências*, 14(2), 151-162. Recuperado de <https://ienci.if.ufrgs.br/index.php/ienci/article/view/353>
- Lu, C., Liu, Y., Xu, S., Zhou, S., Mei, H., Zhang, X., ... & Bao, L. (2023). Conceptual framework assessment of knowledge integration in student learning of measurement uncertainty. *Physical Review Physics Education Research*, 19(2), 020145. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020145>
- Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, 18(8), 955-968. <https://doi.org/10.1080/0950069960180807>
- Lubben, F., Allie, S., & Buffler, A. (2010). Experimental work in science. In M. Rollnick (Ed.). *Identifying Potential for Equitable Access to Tertiary Level Science* (pp.135-152). Berlin, Germany: Springer-Verlag.

- Lubben, F., Campbell, B., Buffler, A., & Allie, S. (2001). Point and set reasoning in practical science measurement by entering university freshmen. *Science Education*, 85(4), 311-327. <https://doi.org/10.1002/sce.1012>
- Millar, R., Lubben, F., Got, R., & Duggan, S. (1994). Investigating in the school science laboratory: conceptual and procedural knowledge and their influence on performance. *Research Papers in Education*, 9(2), 207-248. <https://doi.org/10.1080/0267152940090205>
- MEC - Ministério da Educação (2018). *Base Nacional Comum Curricular*. Recuperado de <http://basenacionalcomum.mec.gov.br>
- Moreira, M. A. (2011). *Metodologias de pesquisa em ensino*. São Paulo, SP: Livraria da Física.
- NRC - National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, United States of America: National Academies Press. Recuperado de <https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts>
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, United States of America: National Academies Press. Recuperado de <https://nap.nationalacademies.org/catalog/18290/next-generation-science-standards-for-states-by-states>
- OCDE. *PISA 2018 assessment and analytical framework*. Paris: OECD publishing, 2019. Recuperado de <https://www.oecd.org/education/pisa-2018-assessment-and-analytical-framework-b25efab8-en.htm>
- Osborne, J. (2016). Defining a knowledge base for reasoning in science: The role of Procedural and Epistemic knowledge. In R. A. Duschl, & A. S. Bismack (Eds.). *Reconceptualizing STEM education: The central role of practices* (pp. 215-231). London, United Kingdom: Routledge.
- Pigosso, L. T., & Heidemann, L. A. (2023). Uma revisão da literatura sobre a abordagem do processo de medição científica no ensino de Física na Educação Básica. *Investigações em Ensino de Ciências*, 28(2), 332–351. <https://doi.org/10.22600/1518-8795.ienci2023v28n2p332>
- Pollard, B., Werth, A., Hobbs, R., & Lewandowski, H. J. (2020). Impact of a course transformation on students' reasoning about measurement uncertainty. *Physical Review Physics Education Research*, 16(2), 020160. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020160>
- Pols, C. F. J., Dekkers, P. J. J. M., & de Vries, M. J. (2022). Defining and assessing understandings of evidence with the assessment rubric for physics inquiry: Towards integration of argumentation and inquiry. *Physical Review Physics Education Research*, 18(1), 010111. <https://doi.org/10.1103/PhysRevPhysEducRes.18.010111>
- Rollnick, M., Dlamini, B., Lotz, S., & Lubben, F. (2001). Views of South African chemistry students in university bridging programs on the reliability of experimental data. *Research in Science Education*, 31, 553-573. <https://doi.org/10.1023/A:1013102108541>
- Rollnick, M., Lubben, F., Lotz, S., & Dlamini, B. (2002). What do underprepared students learn about measurement from introductory laboratory work? *Research in Science Education*, 32, 1-18. <https://doi.org/10.1023/A:1015022804590>
- Sasseron, L. H., & Carvalho, A. M. P. (2011). Alfabetização científica: uma revisão bibliográfica. *Investigações em Ensino de Ciências*, 16(1), 59-77. Recuperado de: <https://ienci.if.ufrgs.br/index.php/ienci/article/view/246>
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32(1), 102. <https://doi.org/10.1037/0012-1649.32.1.102>
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland (KMK). (2020). *Bildungsstandards im Fach Physik für die Allgemeine Hochschulreife*. Berlin, Deutschland: KMK. Recuperado de

https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Physik.pdf

Séré, M. G., Journeaux, R., & Larcher, C. (1993). Learning the statistical analysis of measurement errors. *International Journal of Science Education*, 15(4), 427-438. <https://doi.org/10.1080/0950069930150406>

Varelas, M. (1997). Third and fourth graders' conceptions of repeated trials and best representatives in science experiments. *Journal of Research in Science Teaching*, 34(9), 853-872. [https://doi.org/10.1002/\(SICI\)1098-2736\(199711\)34:9<853::AID-TEA2>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2736(199711)34:9<853::AID-TEA2>3.0.CO;2-T)

Recebido em: 05.04.2024

Aceito em: 20.08.2024