International Society of the Learning Sciences

# Students Reasoning About Bias and Ethics When Designing Human Subjects Research Studies

Camillia Matuk, New York University, cmatuk@nyu.edu
Bernice d'Anjou, New York University, bed6704@nyu.edu
Pranali Mansukhani, New York University, pm3598@nyu.edu
Julie Rissman, Dubin Research & Consulting, jlr9752@nyu.edu
Felicia Zerwas, New York University, fz2338@nyu.edu
Franck Porteous, New York University, fp2221@nyu.edu
Kim Burgas, Independent, burgaska@gmail.com
Yury Shevchenko, University of Konstanz, shevchenko_yury@mail.ru
Suzanne Dikker, New York University, suzanne.dikker@nyu.edu

**Abstract:** Understanding bias and ethics in research is critical for both researchers and consumers of research. It facilitates sound decision-making, builds public trust in research, and ensures that research contributes positively to society. Authentic research opportunities can be critical for developing these skills, but are lacking in high school contexts. We designed and implemented MindHive, a curriculum and online platform for students to design and conduct online human behavior studies. Qualitative analysis of 98 student-generated research proposals and 21 individual interviews with students from across 6 high schools, identified various ways that students engaged with bias and ethical issues within their projects. Engaging in research about personally relevant questions and contexts may support students in shuttling between personal experiences and data, and taking on perspectives that help to reveal and address assumptions. Findings contribute to an understanding of how curriculum can meaningfully engage students in reasoning about research bias and ethics.

## Introduction

Research literacy is not only important for researchers, but also for consumers of research to be able to make informed, evidence-based decisions (Yacoubian, 2018). Understanding research involves understanding the mechanics of the research cycle, including how to ask answerable questions, define and operationalize variables, test hypotheses through methods that appropriately vary and control variables, and draw appropriate conclusions based on evidence. These skills cross-cut domains, and have been variously described in the literature on science literacy (National Academies of Sciences, Engineering, and Medicine, 2016), information literacy (Jones-Jan et al., 2021), statistical literacy (Watson & Callingham, 2003), evidence-based reasoning (e.g., Brown et al., 2010), and critical thinking (Dunn et al., 2008).

However, numerous challenges with understanding research have been documented among middle school, high school, and university level students (Roche et al., 2020). For instance, students struggle to understand and apply the concepts and skills of research, such as articulating researchable hypotheses, identifying and operationalizing variables, selecting methods that properly control and vary factors, and inferring appropriate conclusions based on evidence (Dasgupta et al., 2014). They also harbor feelings of anxiety toward, and disinterest in research, as well as perceptions of its irrelevance (Earley, 2014; Woolley et al., 2018).

More recent conceptions of these literacies additionally emphasize the critical, ethical and moral considerations in the conduct and application of research to real-world problems (Halpern & Dunn, 2021; Kilag et al., 2023; Villadares, 2021; Weiland & Sundrani, 2022). Indeed, two important and related skills in research are in managing bias, and reasoning about ethics. Bias can threaten the validity of a research study, and in turn, have societal implications for how that research is applied in practice.

Hands-on experiences with research can be effective for developing learners' research literacy, and especially for grappling with ideas about research bias and ethics (Kirshner et al., 2011). When given agency to ask, and to conceptualize the manner in which to answer questions that are relevant to them, students can not only develop more realistic notions of the social nature of research, but also come to envision themselves as capable of contributing meaningfully to the generation of scientific knowledge (Harris et al., 2020; Robnett et al., 2015; Buchanan, 2019). At the same time, authentic research experiences can be difficult to facilitate in classrooms that lack necessary time and resources, and adequate support for teachers (Fitzgerald et al., 2019).

This study is part of a larger effort to develop a web-based platform for conducting authentic human subjects research. Through an accompanying high school curriculum, the project aims to address the cognitive,

logistic, and emotional barriers to student-led inquiry. Here, we explore how students come to reason about bias and ethics when designing their own human subjects research studies.

## Theoretical framework

### Managing bias

Bias can come from various sources, and at all stages of a research process. Examples of bias include selection bias (e.g., when members of a population are not represented in a sample due to, for instance, refusal to participate, attrition, or being less likely to volunteer); information bias (when variables are improperly measured or recorded due to, for instance, participants' abilities to recall certain events better than others, or participants' and/or researchers' awareness of a study's hypothesis, or experimental condition to which a participant is allocated); cognitive bias (when responses are affected by, for example, social influence, motivation, or emotions) (Webb et al., 2001; Campbell & Russo, 2001). Researchers must learn to manage bias in designing and conducting their studies; and in analyzing, interpreting, and presenting findings from their studies. For consumers of research, evaluating the validity and reliability of data, and thus, inferring appropriate conclusions based on those data, also requires understanding the various ways that bias can come into play. When not properly managed, bias can lead to wrongly informed decisions that may harm communities they intend to help, and ultimately erode public trust in science.

Drawing on traditions in scientific reasoning and hermeneutics, as outlined by Kirshner et al. (2016), we conceptualize the management of bias as a dynamic process of revisiting and revising assumptions based on critical examinations of personal experiences, and re-evaluations of data. Training to become aware of the various kinds of bias to which people are susceptible can help researchers to avoid bias (Poos et al., 2017). Additionally, to recognize bias, researchers can step away from their hypotheses, and view their work from multiple different perspectives, including those of their participants.

### Ethical reasoning

Ethical reasoning in research requires consideration of its broader societal implications (Steinmann et al., 2016). We draw on a set of ethical principles shared by professionals in research, data science and medicine, and that include *non-maleficence, beneficence, justice, autonomy,* and *trust*.

*Nonmaleficence*, is the principle to 'do no harm,' where harm in behavioral research is most likely to come from violation of privacy, or psychological distress. *Beneficence*, or, to 'do good', is not strictly an obligation of researchers, but is often used to justify the reasons for research. Considered as care for others' well-being, beneficence should factor into reasoning about the tradeoffs of research that may have benefit to some, but cause harm in other unintended ways. *Justice* relates to how opportunities are distributed to groups as a result of research. *Justice* becomes relevant when findings from research, or research that excludes disadvantaged groups, can lead to certain groups being treated differently, such as through discrimination, or unequal distribution of resources. *Autonomy* is the principle of respecting the freedom of participants to make decisions, and ensuring that information is disclosed (such as through informed consent) that allows participants to make informed decisions. *Trust* ensures confidence on the part of participants, that researchers will honor their autonomy, and act with principles of nonmaleficence and beneficence in mind. Adhering to principles of trust involves researchers' transparency about their intentions, and enables them to take risks to explore novel and innovative ideas (Steinmann et al., 2016).

Importantly, these principles are not discrete, but often dependent on one another. For example, trust requires honoring participant autonomy, and beneficence requires considerations of justice and nonmaleficence, and vice versa. Certain principles also gain primacy over others, such as when a researcher suspects abuse or neglect of a minor and are mandated to breach confidentiality to report it, they put nonmaleficence and beneficence over trust and autonomy. Similarly, the goals of research can sometimes justify tradeoffs, such as when a study requires violation of participant autonomy through deception in order to carry out research that may contribute to beneficence or justice.

To negotiate these decisions requires researchers to shuttle between their own research goals, and participants' perspectives. Such reasoning about bias and ethics is facilitated by social interaction, and by curriculum structures that offer multiple opportunities to reveal biases, explain one's thinking, and to view and re-frame the purpose and potential impacts of one's research (Kirshner et al., 2011). With such support in place, individuals can verify their points of view against those of others, recognize perspectives they may not have considered, and receive feedback that encourages them to surface and reconsider their assumptions (Philips & Burbules, 2000).

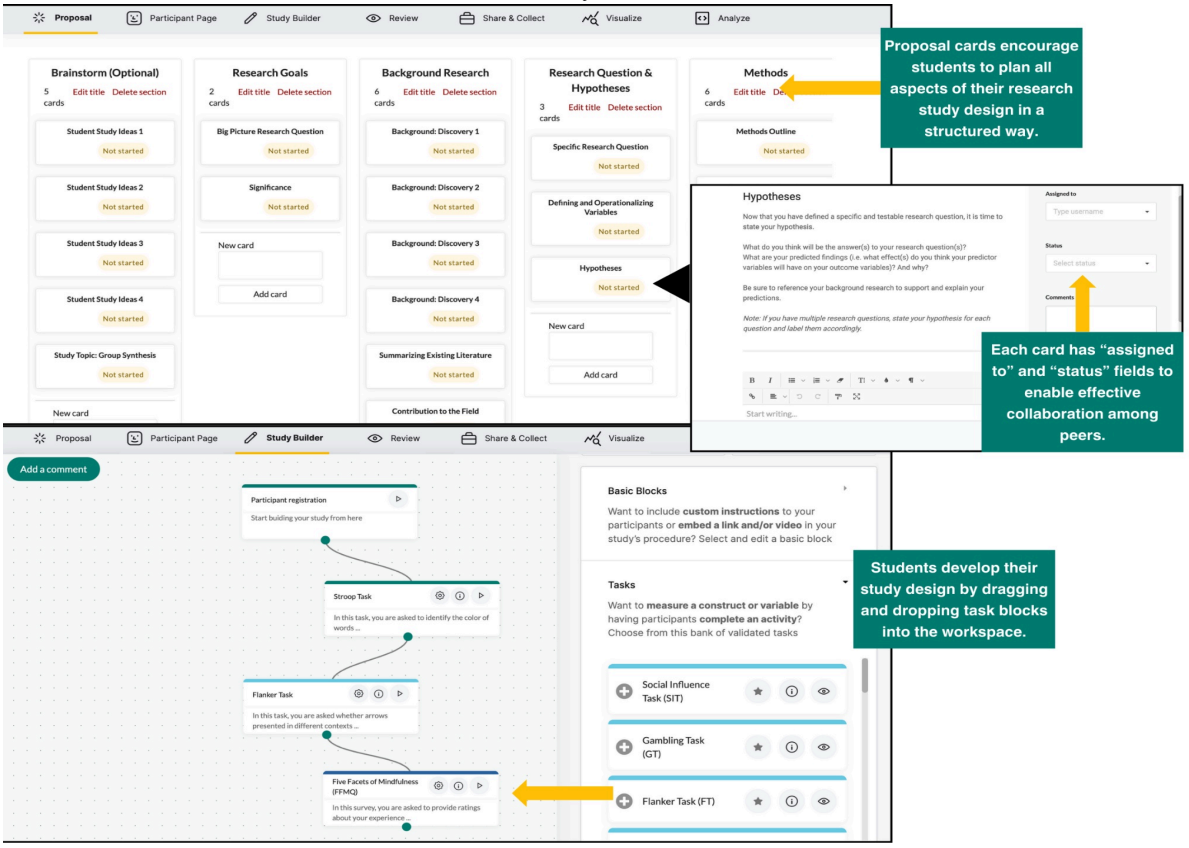
This study asks: How do students reason about bias and ethics in research when designing their own human subjects research? To answer this question, we use the frameworks described above to guide the design of a high school research methods curriculum, and to identify and describe instances of students reasoning about bias and ethics in their projects.

## Methods

### The MindHive platform

This study focuses on students' learning on MindHive (https://mindhive.science), a web-based platform and high school curriculum for human behavior research (Dikker et al., 2022; Matuk et al., 2023, 2024). Various tools on the platform (see Figure 1) enable students to explore and participate in studies created by other users; to develop proposals for their own research studies by completing and arranging drag-and-drop components (e.g., "research question," "participants," "procedure"); to design studies by creating and arranging content, including instructions and weblinks, and tasks and surveys that may be created by students, or adapted from the validated tasks and surveys available in MindHive's repository. For example, students might test how music tempo influences selective attention by creating a study that presents participants with audio excerpts and a Stroop Task. Other tools on MindHive enable students to exchange peer reviews with other students; and to collect and analyze data collected from their study participants. Meanwhile, teacher-facing tools enable the creation of class networks for peer review, and for tracking and giving feedback on students' progress.

**Figure 1**
*The MindHive Platform*




Screenshots from MindHive showing the Proposal Board (top) and StudyBuilder (bottom) used to guide students' design of research studies.

### The MindHive curriculum

Each teacher received access to the MindHive platform with a corresponding semester-long curriculum comprising two units, which they adapted as necessary to fit the goals of their courses. The first unit focuses on the student-as-learner/participant by providing students with the background necessary to understand the

research process, and experience with being a research participant. This unit grounds research in students' own experiences with an emphasis on brain and behavior research. Students also explore models of studies by participating in featured studies created by researchers and other students. For example, one of the featured studies has participants complete a gambling task to understand whether age relates to risk taking behaviors.

One lesson in this unit introduced research ethics. In that 1-class period long lesson, teachers led a class discussion of historical cases of research done without participant consent, and about the responsibilities of scientists toward society. Students also generated questions that can be answered through research on human subjects, and plans for addressing ethical issues surrounding those questions. Part of this activity involved students completing an adapted version of a Responsible Conduct of Research course, through which they become familiar with protocols for data privacy related to research data's purpose, use, openness, access, and security.

Finally, the second unit focused on the student-as-scientist. It guided students in creating their own research projects, from generating ideas, to articulating research questions and hypotheses, to motivating the need for their study, to designing their methods, to building and deploying their study to collect data (from peers in their schools), to visualizing and communicating their findings. Scaffolds on the MindHive platform guided students in structuring and elaborating their proposals, with prompts for students to articulate such components as Research Question, Defining and Operationalizing Variables, Population and Sample, Procedure, Participant Experience and Ethical Considerations, Potential Confounds, Alternative Outcomes, Study Limitations, and Future Directions. Meanwhile, feedback from teachers, and scaffolded peer reviews at key stages of their projects, supported students' iterative refinement of their study designs.

## Participants and context

A total of 237 students and 7 teachers used MindHive during the 2022-2023 academic year over the course of 10 weeks. Students completed their MindHive projects individually, or in groups of 2-4 students. Participants were from 6 high schools (2 private, 4 public) in a large, urban, socioeconomically and culturally diverse city in the northeastern United States. MindHive was implemented in a total of 11 class sections of 7 courses, with some of the 7 teachers teaching multiple sections of the same course. Two of the courses were Environmental Science electives (1 course in grade 9, 1 course in grades 11-12). Five courses were Neuroscience electives (1 course in grades 10-12, 1 in grade 11, and 3 in grades 11-12). All students were using MindHive for the first time during this implementation. Teachers ranged in their prior experiences with MindHive, with some having taught with it for two years previously, and others teaching it for the first time during this implementation. To support teachers during their curriculum implementation, our team offered weekly virtual office hours, during which teachers brought issues to troubleshoot with our designers and developers, and ideas and instructional strategies to share and discuss with fellow teachers.

## Data and analysis

Our data come from individual interviews with 21 students (selected based on obtained consent), across 5 schools. Students from one of the public schools were not interviewed due to their limited class time. Interviews were 30-minutes long, and asked students to describe their study's design and motivation (e.g., What methods did you choose to answer this question? Why?). We also collected the 98 proposals created in MindHive by students across 5 schools. One of the public school teachers preferred to complete the proposals outside of MindHive, and so we excluded their students' work from our analysis.

Through rounds of independent, inductive coding and discussion of the interview transcripts, three researchers identified emergent themes (Strauss & Corbin, 1998) related to students' reasoning about bias and ethics. Through discussion with a fourth researcher, they reached consensus on a set of codes, which the three first researchers then applied in a deductive thematic analysis (Braun & Clark, 2006) of the research proposals. Codes were applied based on presence or absence of a theme, looking holistically at the proposal (a single one of which ranged from ~540-1940 words) as a unit of analysis, and resolving disagreements through discussion. After rounds of independent coding and discussion on subsets of the proposals, a final set of codes was established with substantial interrater agreement across categories (Cohen's Kappa = 0.66-1, see Table 1). One researcher then coded the remaining proposals. We report descriptive statistics for each theme to demonstrate their prevalence across students' research proposals.

# Results

The research questions of students' proposals ranged widely, with some asking how music tempo impacts risk-aversion; how social perceptions affect a person's sense of confidence; how empathy relates to people's sense of care toward animals; and how spatial awareness differs between neurotypical and neurodivergent

individuals. Themes that emerged from our coding of students' interviews and proposals fell into two categories: One related to students' reasoning about sources of bias in their study designs, and a second related to ethical considerations in the design and conduct of their research (Table 1). Below, we describe sub themes of these categories, and illustrative examples from students' work.

**Table 1**
*Percentage of Student Proposals Meeting Subtheme Criteria*

| | % of Proposals Meeting Criteria | Cohen's Kappa |
|---|---|---|
| **Sources of Bias** | | |
| Response bias | | |
| Survey fatigue | 39% | 0.86 |
| Mundane realism | 45% | 0.78 |
| Selection bias | 34% | 0.66 |
| Information bias | | |
| Measurement bias | 14% | 1 |
| Participants' cognitive biases | 37% | 1 |
| **Ethical considerations** | | |
| Trust / Autonomy | | |
| Informed consent | 20% | 1 |
| Privacy | 28% | 0.93 |
| Nonmaleficence | | |
| Protection of vulnerable populations | 19% | 0.85 |
| Avoidance of sensitive or invasive questions | 27% | 1 |
| Beneficence | 11% | 0.90 |

## Students' reasoning about bias

Students identified various sources of potential bias in reflecting on their research studies, including (non)response bias, selection bias, and information bias.

Non-response bias—when respondents fail to give responses, leading to missing data—tended to be identified in the recognition of potential *survey fatigue* (39% of proposals). More specifically, students who reflected on non-response bias noted how the lengths of their surveys may have failed to align with participants' attention span and interest, and so impacted their overall willingness to complete those surveys. These students noted how such issues could lead to incomplete data that would affect the reliability of conclusions. For example, in their proposal to study perceived peer performance on confidence, Student 5 wrote:

> *"Ideally we would be able to have more tests because it would allow for more opportunities to negatively or positively reinforce the participant. However, this would cause participant fatigue which could lead to worse actual results"*

Student 16 provided a nonexample when they identified how their study included a long task, but did not comment on how the length would affect results. *"I learned how other people felt when they were doing their study. That [it] was kind of long."*

Meanwhile, response bias—when respondents give inaccurate or untruthful responses—was identified by students in their attention to principles of *Mundane realism* (45% of proposals), that is, the degree to which research study tasks and questions reflect participants' actual experiences, and so are more likely to elicit responses that participants would realistically give (Kelly, 2007). For example, in their study examining the relationship between people's opinions of animals and their empathy levels, Student 11 wrote:

> *"We wanted to create a survey that includes a visceral reaction of how people see these mammals with photos instead of just naming the animal and having them recall it from memory because we felt that the study would be more accurate if it wasn't all personal opinion."*

In contrast, Student 16 neglected to consider the impacts of mundane realism on response bias when they rationalized their study's questionnaire by simply describing:

> *"Our study is reasonable because it is not too demanding and the instructions are clear. The participants will be able to complete the survey easily because the questions are straightforward and clear. The questions will have 7 options ranging from strongly disagree to strongly agree."*

*Selection bias* (34% of proposals) was a second source of bias that students identified. This was observed when students reflected on how characteristics of their participants might introduce confounds. Student 5 provided a representative example in the context of participants' performance on logic tests in their study about the effects of perceived peer performance on confidence. *"People who have better baseline skill will likely have an innately higher confidence and perform better.*" In contrast, Student 16 provided a nonexample by focusing more on the misalignment of the participant group with the study goal: *"But it might be just because we did [it] in a small like participant group and then you might most because of them were students, who you know, have not really been to casinos so."*

A third source of bias that students identified was *Information bias*. This appeared in students' reflections on two different kinds of information bias. One was *Measurement bias* (14% of proposals), that is, when variables are measured incorrectly. For example, in their study measuring the impact of music on adolescents' attention spans, Student 4 described how environmental conditions may have interfered with participants' abilities to hear the music as intended in their treatment conditions. They wrote:

> *"[A] potential confound is we cannot control the environment that the participants are doing the experiment in. Therefore, there could be external distractions and noises that could conflict with the music they are listening to. In addition, the type of headphones that each of the participants is wearing could be a potential confound because some could be noise canceling ensuring they cannot have any external noise distractions, whereas others may be able to hear other distractions."*

In contrast, Student 19, whose study investigated how people of different personality types respond to stress, pointed out *Selection and Sampling bias*, but neglected to consider how the physical environment (e.g., noise in the classroom where the study took place) might have contributed to participants' stress responses:

> *"Some potential confounding variables in this study could be gender, socioeconomic status, and prior experience with stress. These variables could influence participants' stress response and mindfulness levels and therefore could impact the study results."*

A second kind of information bias that students identified was in *participants' cognitive biases* (37% of proposals), and specifically, the experiences, attitudes and beliefs that may influence their responses. For example, in their study measuring the effect of owning a pet on happiness, Student 7 wrote: *"Participants could have recently lost a pet. This would associate pets with sadness, therefore, skewing their data to be more unhappy."*

## Students' reasoning about ethical considerations

Students also discussed the various ways that they considered ethical issues in their research studies, particularly under the ethical principles of *trust and autonomy,* and *nonmaleficence*.

In terms of *Trust and Autonomy*, students reflected on the importance of *Informed Consent* studies (20% of proposals). For instance, in their study exploring how different personality types respond to stress, Student 19 wrote of the importance of offering potential participants a full picture of the study's potential harms:

> *"... our study involves stress-inducing tasks that may be potentially harmful to participants. As a precautionary measure, we will provide participants with a disclaimer and clear terms and conditions outlining the study's nature and potential risks. We will also take steps to ensure that participants fully understand the study's purpose and potential risks before agreeing to participate."*

Also related to the principles of *Trust and Autonomy* was students' attention to protecting participant *Privacy* (28% of proposals). For example, Student 5, examining the relationship between perceived peer performance and self-confidence, wrote:

> *"The only questions that pertain to the participants themselves will be asked in the demographic survey that participants will take after participating in the main part of the study. Participation will be anonymous for each participant."*

Students' reasoning about trust and autonomy also showed in their awareness of the principle of *Nonmaleficence*, that is, in the need to avoid harm that could be done should trust and autonomy be breached. This appeared in some students' attention to *Protecting vulnerable populations* (19% of proposals); their *Avoidance of sensitive or invasive questions* (27% of proposals); and their instinct to ensure the *Beneficence* of their research (11% of proposals).

For example, in their study exploring the relationship between facial structure and attractiveness, Student 22 wrote:

> "*We are studying a vulnerable population because we will be conducting it on a high school group and most of us are teens that are adapting and won't feel comfortable."* They furthermore explained: "*We will take measures to make everything anonymous and prevent any form of bullying or harassment by taking away their phones before experimenting so people are comfortable not getting harassed."*

Meanwhile, Student 7, whose study examined the effect of owning a pet on happiness, described their intentional avoidance of tasks or questions that may be too personal: *"The questions will not really be sensitive or invasive. The answers will be anonymous. If there is a question that is particularly sensitive or hard to answer, we will include an option 'prefer not to answer'."*

In terms of the principles of *Beneficence* and *Nonmaleficence*, students reasoned about their own responsibilities as researchers to contribute positively to society, and to avoid exacerbating the issues that they are studying. For instance, in their study on the effects of ADHD on spatial awareness, Student 2 wrote:

> *"This could be seen as us having a negative assumption of a neurodivergent group by hypothesizing that the neurodivergent group, as a minority in the world, is less capable of higher levels of spatial awareness as defined by our study. However, this study merely aims to form conclusions on spatial perception and awareness, and not the capabilities of people with ADHD as a whole."*

Meanwhile, Student 22, studying the relationship between facial structure and attractiveness, provided a contrasting case when they reasoned about their participants' role in exacerbating issues related to social perceptions of attractiveness, but neglected to consider their own responsibility in engaging in this study to : *"It may result in a negative way, since not everyone thinks the same, maybe some will feel insecure. Or maybe they will judge based on the cleanliness of their face, it all depends on the factors they judge by."*

## Discussion and Conclusion

Students in this study demonstrated abilities to reason about different kinds of bias and ethical issues in their research studies. While some categories of bias and ethics appear to have been more frequently mentioned by students than others (e.g., *Mundane realism* at 45% vs. *Beneficence* at 11%), it is possible that these were simply not equally salient across the various types of studies that students chose to pursue. For example, ethical issues are likely more apparent in a study about perceptions of attractiveness, than in a study about the impacts of music tempo on risk aversion.

It is also possible that students' own knowledge and experiences limited them to notice certain kinds of bias and issues of ethics. Notably, the kinds of bias that students identified appeared to largely relate to characteristics known by student-researchers of their participant population (i.e., their school peers). It may be that, because students shared the personal experiences of their participants, and because their studies explored questions about issues that directly concerned them, students brought personal perspectives to their research that enabled them to easily identify certain issues of bias and ethics. In other words, student-researchers could personally relate to their population's experiences of fatigue from too-long surveys, the sadness of having recently lost a pet, the sense of insecurity at the thought of being judged by peers, which they also identified as issues of bias and ethics. This possibility aligns with Kirshner et al.'s (2016) assertion that managing bias and

the ethical issues that bias entails is accomplished by shuttling between personal experiences and data in order to surface and address hidden assumptions.

Our ongoing research might examine the contributions of individual teachers, classroom-specific activities, and of the MindHive curriculum, to students' learning to reason about bias and ethics in research. Future research might also explore how curriculum and instruction might be designed to support students in reasoning about the many other forms of bias and ethical considerations that students in our sample did not reflect upon; and on the necessary support for teachers to effectively facilitate these learning experiences (Fitzgerald, 2019).

Overall, this study suggests that students, given the opportunity to design and conduct their own research, can come to recognize and reason about sources of bias. Importantly, some students reasoned about these issues in terms of how they may have impacted the quality of data, and thus, the conclusions that can be drawn from those data—a skill that is essential for evaluating research, and making sound decisions. Such authentic research experiences can also offer contexts in which students can reason about ethical considerations in the design and conduct of research, and on the responsibilities of researchers toward society. These understandings are critical for building public trust in research, and for encouraging exploration and innovations that may benefit society (Steinmann et al., 2016).

## References


Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Brown, N. J., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*(3-4), 123-141.

Buchanan, S. C. (2019). Using the hermeneutic phenomenological method to explore the middle school student lived experience of student driven inquiry. *Qualitative and Quantitative Methods in Libraries, Special Issue: School Library Research and Educational Resources, 6*, 61–74.

Campbell D. T., Russo M. J. (Eds.). (2001). *Social measurement*. Thousand Oaks, CA: SAGE.

Dasgupta, A. P., Anderson, T. R., & Pelaez, N. (2014). Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties. *CBE—Life Sciences Education, 13*(2), 265-284.

Dikker, S., Shevchenko, Y., Burgas, K., Chaloner, K., Sole, M., Yetman-Michaelson, L., ... & Matuk, C. (2022). MindHive: An online citizen science tool and curriculum for human brain and behavior research. *Connected Science Learning*, *4*(2), 12318663.

Dunn, D. S., Halonen, J. S., & Smith, R. A. (Eds.). (2008). *Teaching critical thinking in psychology: A handbook of best practices*. Malden, MA: Wiley

Earley, M. A. (2014). A synthesis of the literature on research methods education. *Teaching in Higher Education*, *19*(3), 242–253.

Fitzgerald, M., Danaia, L., & McKinnon, D. H. (2019). Barriers inhibiting inquiry-based science teaching and potential solutions: Perceptions of positively inclined early adopters. *Research in Science Education, 49*(2), 543–566.

Halpern, D. F., & Dunn, D. S. (2021). Critical thinking: A model of intelligence for solving real-world problems. *Journal of Intelligence, 9*(2), 22.

Harris, E. M., Dixon, C. G. H., Bird, E. B., & Ballard, H. L. (2020). For science and self: youth interactions with data in community and citizen science. *Journal of the Learning Sciences*, *29*(2), 224–263.

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, *65*(2), 371-388.

Kelly J. (2007). Mundane realism. In R. F Baumeister, K. D Vohs (Eds.), *Encyclopedia of Social Psychology* (Vol. 1, pp. 599599). SAGE Publications, Inc.

Kilag, O. K., Lisao, C., Lastimoso, J., Villa, F. L., & Miñoza, C. A. (2023). Bildung-Oriented Science Education: A Critical Review of Different Visions of Scientific Literacy. *Excellencia: International Multi-disciplinary Journal of Education*, *1*(4), 115-127.

Kirshner, B., Pozzoboni, K., & Jones, H. (2011). Learning how to manage bias: A case study of youth participatory action research. *Applied Developmental Science*, *15*(3), 140-155.

Matuk, C., d'Anjou, B., Mansukhani, P., Porteous, F., Burgas, K., Zerwas, F., ... & Dikker, S. (2024, June). Tools to Support High School Students' Creativity in Scientific Research. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (pp. 655-659).

Matuk, C., Yetman-Michaelson, L., Martin, R., Vasudevan, V., Burgas, K., Davidesco, I., ... & Dikker, S. (2023).

Open science in the classroom: students designing and peer reviewing studies in human brain and behavior research. *Instructional Science*, *51*(5), 793-845.

National Academies of Sciences, Engineering, and Medicine. (2016). *Science literacy: Concepts, contexts, and consequences*. Washington, DC: The National Academies Press. doi:10.17226/23595.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Poos, J. M., van den Bosch, K., & Janssen, C. P. (2017). Battling bias: Effects of training and training context. *Computers & Education*, *111*, 101-113.

Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching, 52*(6), 847–867. https://doi.org/10.1002/tea.21221

Roche, J., Bell, L., Galvão, C., Golumbic, Y. N., Kloetzer, L., Knoben, N., ... & Winter, S. (2020). Citizen science, education, and learning: Challenges and opportunities. *Frontiers in Sociology*, *5*, 613814.

Steinmann, M., Matei, S. A., & Collmann, J. (2016). A theoretical framework for ethical reflection in big data research. In, J. Collmann & S. A. Matei (Eds.), *Ethical reasoning in big data: An exploratory analysis (pp. 11-27)*. Springer.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Valladares, L. (2021). Scientific literacy and social transformation: Critical perspectives about science participation and emancipation. *Science & Education*, *30*(3), 557-587.

Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46.

Webb E. J., Campbell D. T., Schwartz R. D., Sechrest L. B., Grove J. B. (2001). Fifteen validity issues affecting all measures. In, Campbell D. T. & Russo M. J. (Eds.), *Social measurement* (pp. 137-171). Thousand Oaks, CA: SAGE.

Weiland, T., & Sundrani, A. (2022). Towards a framework for developing a critical statistical literacy. In, S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), Bridging the Gap: Empowering & Educating Today's Learners in Statistics. *Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022)*, Rosario, Argentina. International Association for Statistical Education.

Woolley, J. S., Deal, A. M., Green, J., Hathenbruck, F., Kurtz, S. A., Park, T. K. H., Pollock, S. V., Bryant Transtrum, M., & Jensen, J. L. (2018). Undergraduate students demonstrate common false scientific reasoning strategies. *In Thinking Skills and Creativity, 27*, 101–113. https://doi.org/10.1016/j.tsc.2017.12.004

Yacoubian, H. A. (2018). Scientific literacy for democratic decision-making. *International Journal of Science Education*, *40*(3), 308-327.

## Acknowledgments

This research was supported by awards from NSF (1908482, 2241751) and NIH SEPA (1R25MH135446-01); and by Dana Bevilacqua, Monica Beneyto, Patrick Bloniasz, Kim Chaloner, Girija Chatufale, Jennifer Dahlstrom, Andrew Ferreira, Xavier Ochoa, Deborah Reich, Esteban Romero, and Marc Sole.