Educational Technology & Society has Impact Factor 2.633 and
5-Year impact factor 4.358 according to Thomson Scientific 2021 Journal Citations Report.

http://www.j-ets.net/

Oct. 2022

Educational Technology & Society

vol.25
no.4

Educational Technology & Society

Published by International Forum of Educational Technology & Society
Hosted by National Taiwan Normal University, Taiwan

Volume 25 Issue 4

October 2022

ISSN: 1436-4522 (online)
ISSN: 1176-3647 (print)
http://www.j-ets.net/

Educational Technology & Society
An International Journal

Aims and Scope

Educational Technology & Society (ET&S) is an open-access academic journal published quarterly (January, April, July, and October) since October 1998. *ET&S* has achieved its purposes of providing an international forum for open access scientific dialogue for developers, educators and researchers to foster the development of research in educational technology. Thanks to all the Authors, Reviewers and Readers, the journal has enjoyed tremendous success.

ET&S has established a solid and stable editorial office where the Editors-in-Chief aims to promote innovative educational technology research based on empirical inquiries to echo the pedagogical essentials of learning in the real world—lifelong learning, competency-orientation, and multimodal literacy in the 21st century.

ET&S publishes research that well bridges the pedagogy and practice in advanced technology for evidence-based and meaningful educational application. The focus of *ET&S* is not only technology per se, but rather issues related to the process continuum of learning, teaching, and assessment and how they are affected or enhanced using technologies. The empirical research about how technology can be used to overcome the existing problems in the frontline of local education with findings that can be applied to the global spectrum is also welcome. However, papers with only descriptions of the results obtained from self-report surveys without systematic or empirical data or any analysis on learning outcomes or processes are not favorable to be included in *ET&S.*

ET&S publishes research that fulfills any of the following description:

- **Evidence-based Research:** Continuous research that is conducted within a sufficient amount of time such that the effectiveness of the research intervention in education can be evaluated and demonstrated convincingly via empirical methods. Preferably, research experience, outcome, and results for a semester long of adopting educational technologies; or an evaluation/experiment that is conducted multiple times and spans over several semesters.
- **Emerging Technology:** New and emerging technology used in education
- **Special Domain Research:** Research with specific participants using technology in education, for example, those with privacy issues in special education or those of younger ages like infants and toddlers, who may not form a large pool of participants

Publisher
International Forum of Educational Technology & Society

Host
National Taiwan Normal University, Taiwan

Editorial Office
c/o Chair Professor Nian-Shing Chen, Institute for Research Excellence in Learning Sciences, Program of Learning Sciences, National Taiwan Normal University, No.162, Sec. 1, Heping E. Rd., Da-an Dist., Taipei City 10610, Taiwan.

Supporting Organizations
University of North Texas, USA
University of Piraeus, Greece
National Yunlin University of Science and Technology, Taiwan
National Sun Yat-sen University, Taiwan

Website Maintenance
Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan

Advertisements
Educational Technology & Society accepts advertisement of products and services of direct interest and usefulness to the readers of the journal, those involved in education and educational technology. Contact the editors at ets.editors@gmail.com

Abstracting and Indexing
Educational Technology & Society is abstracted/indexed in Social Science Citation Index, Scopus, ACM Guide to Computing Literature, Australian DEST Register of Refereed Journals, Computing Reviews, Current Contents/Social & Behavioral Sciences, DBLP, DOAJ, Educational Administration Abstracts, Educational Research Abstracts, Educational Technology Abstracts, Elsevier Bibliographic Databases, ERIC Clearinghouse on Information & Technology, Inspec, ISI Alerting Services, JSTOR, PsycINFO, Social Scisearch, Technical Education & Training Abstracts, and VOCED.

Guidelines for authors
Submissions are invited in the following categories:

- Peer reviewed publications: Full length articles (up to 8,000 words excluding References and Appendices)
- Special Issue publications

All peer review publications will be refereed in double-blind review process by at least two international reviewers with expertise in the relevant subject area.

For detailed information on how to format your submissions, please see: https://www.j-ets.net/author_guide

For Special Issue Proposal submission, please see: https://www.j-ets.net/journal_info/special-issue-proposals

Submission procedure
All submissions must be uploaded through our online management system (http://www.j-ets.net). Do note that all manuscripts must comply with requirements stated in the Authors Guidelines.

Authors, submitting articles for a particular special issue, should send their submissions according to the channel specified in the Call for Paper of the special issue.

All submissions should be in electronic form. Authors will receive an email acknowledgement of their submission.

The preferred formats for submission are Word document, and not in any other word-processing or desktop-publishing formats. Please place figures and tables in their respective format in the anonymous manuscript along with all appendices (if any).

Please provide following details with each submission in a separate file (i.e., Title Page):

- Author(s) full name(s) including title(s),
- Name of corresponding author,
- Job title(s),
- Organisation(s),
- Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

In case of difficulties, please contact journal.ets@gmail.com.

Educational Technology & Society

Volume 25 Number 4 2022

Table of Contents

Temporal Structures and Sequential Patterns of Self-regulated Learning Behaviors in Problem Solving with an Intelligent Tutoring System

Shan Li*, Juan Zheng and Susanne P. Lajoie
Department of Educational and Counselling Psychology, McGill University, Montreal, QC, Canada
shan.li2@mail.mcgill.ca // juan.zheng@mail.mcgill.ca // susanne.lajoie@mcgill.ca
*Corresponding author

(Submitted July 7, 2021; Revised November 16, 2021; Accepted November 28, 2021)

ABSTRACT: Examining the sequential patterns of self-regulated learning (SRL) behaviors is gaining popularity to understand students' performance differences. However, few studies have looked at the transition probabilities among different SRL behaviors. Moreover, there is a lack of research investigating the temporal structures of students' SRL behaviors (e.g., repetitiveness and predictability) and how they related to students' performance. In this study, 75 students from a top North American university were tasked to diagnose a virtual patient in an intelligent tutoring system. We used recurrence quantification analysis and sequential analysis to analyze the temporal structures and sequential patterns of students' SRL behaviors. We compared the differences between low and high performers. We found that low performers had more single, isolated recurrent behaviors in problem-solving, whereas the recurrent behaviors of high performers were more likely to be part of a behavioral sequence. High performers also demonstrated a higher transition probability across the three phases of SRL than low performers. In addition, high performers were unique in that their behavioral state transitions were cyclically sustained. This study provided researchers with theoretical insights regarding the cyclical nature of SRL. This study has also methodological contributions to the analysis of the temporal structures of SRL behaviors.

Keywords: Self-regulated learning, SRL behavior, Recurrence quantification analysis, Temporal structure, Intelligent tutoring system

1. Introduction

Research in self-regulated learning (SRL) over the past three decades has generated numerous theoretical perspectives on its features, components, phases or subprocesses, and mechanisms, as well as a substantial number of empirical studies (Boekaerts & Niemivirta, 2000; Greene & Azevedo, 2007; Pintrich, 2004; Schunk & Greene, 2017; Winne, 2017; Zimmerman, 2000). Despite the divergent research base, researchers have reached the same conclusion that SRL is essential to high performance in all sorts of learning and problem-solving activities (Broadbent et al., 2020). For this reason, SRL has become a core theoretical framework to understand performance differences among learners across various disciplines and learning contexts. Moreover, researchers generally agree that SRL is a dynamic and cyclical process comprising a series of events that unfold over time in learning tasks (Azevedo, 2014; Panadero, 2017). Therefore, there has been a growing interest in analyzing the features of SRL behaviors, which could provide direct insights on the factors that distinguish self-regulated learners from those lacking SRL competency and the optimal pathway to excellent performance.

Perhaps the most simple and straightforward way to analyze the features of SRL behaviors is through descriptive analysis, which yields a number of descriptive measures such as the total number, frequency, and variance of SRL behaviors. However, descriptive statistics can be misleading if the data is significantly skewed or contains outliers. Descriptive statistics also reduce the information about how SRL behaviors temporally unfold over time. In this study, we introduced an analytical approach, namely recurrence quantification analysis (RQA), to study the temporal structures of SRL behaviors. In particular, RQA was developed to characterize the temporal properties of elements (e.g., repetitiveness and predictability) in event- or time-series data (Wallot, 2017). It provides a number of indices to describe the features of SRL behaviors without losing their richness and temporal information. To our knowledge, no previous attempts have been made to analyze the features of SRL behaviors using the RQA method, which we deem will generate new knowledge about the SRL process.

In addition, researchers are increasingly using advanced analytical methods to uncover the sequential patterns of SRL behaviors in an effort to understand students' SRL strategies and dispositions (Yu et al., 2021). However, there is limited research examining the sequential patterns of SRL behaviors with a focus on the transition probabilities between SRL behaviors. Moreover, the explanation of the SRL behavioral patterns is often confined to a specific research context, leaving little discussion on the theoretical specifications of SRL (e.g., the cyclical nature of SRL) that are represented by those patterns (Paans et al., 2019).

For this study, we take the initiative to demonstrate how RQA can be used to quantify the temporal structures of SRL behaviors and what insights can be brought forth by this method in an empirical study. Moreover, we examine the sequential patterns of SRL behaviors to understand students' SRL strategies and dispositions. This study contributes to the SRL literature in both methodological and theoretical dimensions.

2. Theoretical background

2.1. Self-regulated learning

Self-regulated learning (SRL) is a cyclical process whereby learners purposefully control and monitor their learning or problem-solving strategies in the dimensions of cognition, metacognition, motivation, and emotion, to achieve predetermined goals (Greene & Azevedo, 2007; Li & Lajoie, 2021; Pintrich, 2004; Schunk & Greene, 2017; Winne, 2017; Zimmerman, 2000). Researchers have reached a consensus that SRL processes should be studied as events, which temporally unfold over time during learning and problem-solving (Azevedo, 2014; Chen & Su, 2019; Michailidis et al., 2018). Paans et al. (2019) further argued that temporal variation of SRL occurs at micro- and macro-level time scales, which "develop in parallel and occur at the same time" (p. 247). For example, a certain number of activities at the micro-level (e.g., task analysis, goal setting, and knowledge acquisition) may be cycled and recycled within one macro-level phase, e.g., the planning phase of SRL.

At the micro-level, SRL models describe how learners self-regulate the components of learning, particularly behaviors, emotions, cognitive and metacognitive activities, in a specific learning or problem-solving context. Researchers examined the sequences of those components to understand performance differences among learners. However, there are no strong assumptions on the relationships between specific micro-level sequences of SRL events and task performance (Azevedo, 2014) since SRL at the micro-level is context dependent.

As an illustration, Schoor and Bannert (2012) used process mining to identify sequences of metacognitive activities for high versus low group performance dyads, as students worked in pairs to solve a task related to statistics. They found that there were no major differences in the patterns of regulatory activities between high- and low-performance groups. In another study, Bannert et al. (2014) used the same analytical technique (i.e., process mining) to explore the sequences of students' SRL activities, as they learned specific concepts and principles of operant conditioning. However, Bannert et al. (2014) found that successful students performed SRL activities in a different order when compared with low performers. Bannert et al. (2014) concluded that it would be problematic to compare students' patterns of learning activities in different learning settings, especially when researchers used different types of data and operationalized students' performance differently to extract such patterns.

At the macro-level, the predominant SRL models suggest that successful self-regulated learners generally follow the phases of SRL in time order, although researchers hold different beliefs about the specific phases that consist of SRL process (Greene & Azevedo, 2007; Paans et al., 2019; Pintrich, 2004; Schunk & Greene, 2017; Winne, 2017; Zimmerman, 2000). For example, Winne (2017) proposed a model of SRL that "unfolds over four loosely sequential and recursive phases" (p. 39), i.e., task definition, goal setting and planning, studying tactics, and adaptations to metacognition. In the conceptual framework for SRL, Pintrich (2004) also contended that SRL comprises four phases (i.e., forethought, planning and activation; monitoring; control; and reaction and reflection), which "represent a general time-ordered sequence" (p. 389). According to Zimmerman (2000), students' self-regulatory activities in learning fall in three cyclical phases: forethought, performance, and self-reflection. The forethought phase is followed by the performance phase, which potentially triggers a range of self-reflections that influence forethought in turn. Students may perform many self-regulatory cycles to complete a task.

Nevertheless, there exists little empirical evidence examining how students move through different phases of SRL during a learning task (Paans et al., 2019). Practically speaking, it is recommended to extract the sequential patterns of SRL behaviors at the micro-level, which is a practice that many studies have followed. But we contend that the SRL behavioral patterns should be interpreted at the micro- and macro-levels so that the results can be comparable across different situations, and we can come close to the nature of SRL behaviors.

2.2. Sequential analysis of SRL behaviors

Sequential analysis aims to detect the recurring sequential patterns in a finite stream of actions or events (Gottman et al., 1990). SRL researchers use sequential analysis to extract the patterns of SRL behaviors to gain insights into students' use of SRL strategies and SRL dispositions since the raw sequence of learning activities usually contains redundant and fuzzy information. In addition, SRL behaviors do not always occur in a straightforward and structured manner, and there are significant individual differences in how students perform SRL behaviors. The analysis and interpretation of SRL behavioral patterns provide researchers with a clearer representation of students' cognitive structures and decision-making processes than raw behavioral data. Furthermore, the visualization of SRL behavioral patterns offers an intuitive understanding of the complex, temporally unfolding processes of SRL.

A number of analytical techniques are available in the extant literature to extract the sequential patterns of SRL behaviors, including but not limited to process mining (Schoor & Bannert, 2012), t-pattern analysis (Kuvalja et al., 2014), Hidden Markov Modeling, state-transition analysis, and lag sequential analysis (Bakeman & Quera, 1995; Kapur, 2011). As pointed out by Azevedo (2014), advanced techniques that are capable of analyzing the sequential characteristics of SRL process have the potential to transform contemporary conceptions of SRL.

In this study, we use the lag sequential analysis (LSA) to extract SRL patterns since it has several advantages over other sequential analysis techniques. For one, LSA identifies statistically significant transitions from one type of SRL behavior to another. The transition probabilities between different categories of SRL behaviors can be converted into odds ratios or likelihoods for comparison (Kapur, 2011). More importantly, LSA generates sequential patterns of SRL behaviors that describe the SRL process as a whole since the transition probabilities are calculated for each pair of SRL behaviors. LSA provides a holistic view of an SRL process rather than a number of sub-sequences of events. As aforementioned, the present study complements the ongoing efforts that examine students' SRL behavioral patterns, by introducing the RQA method to depict the temporal structures of SRL behaviors.

2.3. Recurrence quantification analysis

RQA is a nonlinear analysis method that quantifies the dynamics of temporal sequences of change over time by detecting "recurrent events" in a time series (Fleuchaus et al., 2020; Jenkins et al., 2020; Wallot, 2017). When applying RQA on a behavioral time series, it returns a range of RQA metrics. The most commonly used RQA measures are percent recurrence (%*REC*), percent determinism (%*DET*), average diagonal line length (*ADL*), laminarity (%*LAM*), and trapping time (*TT*) (Marwan et al., 2002; Meinecke et al., 2020; Wallot, 2017). To understand these RQA measures, a core concept that needs to be explained is the recurrence plot (RP).

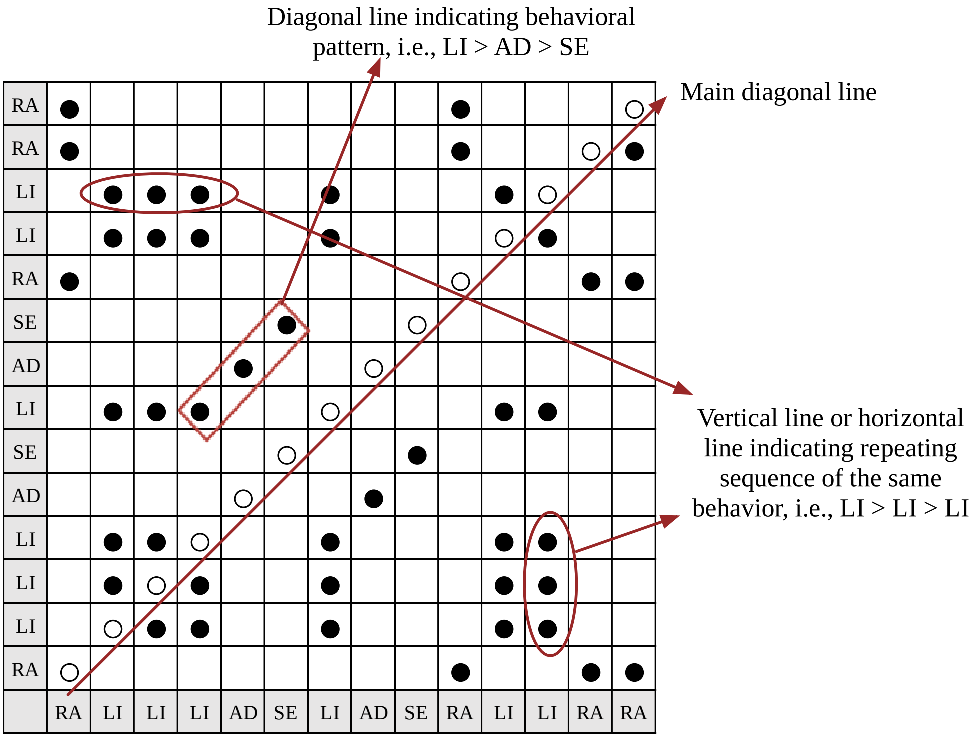
Table 1. The selected RQA measures to quantifying the temporal structure of learning behaviors (Marwan et al., 2002; Meinecke et al., 2020; Wallot, 2017)

Variable	Definition	Meaning
Percent Recurrence (%*REC*)	Percentage of recurrence points in a recurrence plot (RP). %*REC* = Sum of recurrent points in the RP / size of RP	How often does an individual show the same behavior (i.e., repetitiveness of behaviors in the time series)
Percent Determinism (%*DET*)	Proportion of recurrent points forming diagonal lines in a RP. %*DET* = Sum of diagonally adjacent recurrent points / sum of recurrent points	To what extent do repetitions of behaviors occur in the form of behavioral patterns? e.g., a student may conduct a behavior repeatedly. They can also demonstrate certain behavioral patterns.
Average Diagonal Line Length (*ADL*)	Average length of diagonal lines in the RP	How long is the average repeating behavioral pattern?
Laminarity (%*LAM*)	Proportion of recurrent points forming vertical line structures. %*LAM* = Sum of vertically adjacent recurrent points / sum of recurrent points	To what extent do repetitions of behaviors occur in repeating sequences of the same behavior?
Trapping time (*TT*)	Average length of vertical lines in the RP	How long is the average repeating sequence of the same behavior?

RP is the visualization of the recurrence values within a discrete time-series by plotting the time series on both the *x* and *y*-axis of a two-dimensional grid. Figure 1 shows the illustration of a recurrence plot. In RP, the adjacent points that form a vertical or horizontal line signify a repeating sequence of the same behavior, e.g., linking evidence (LI) → linking evidence (LI) → linking evidence (LI). The diagonal lines in RP indicate that students demonstrate a sequential behavioral pattern, for example, linking evidence (LI) → ordering lab test (AD) → searching library (SE). The calculation of RQA measures is based on the distribution of the recurrent points in the RP. For instance, %*REC* equals the percentage of recurrent points in an RP, and %*DET* is the proportion of recurrent points forming diagonal lines in an RP.

In particular, the definition of each RQA measure of interest and its corresponding meaning are shown in Table 1. It is noteworthy that the recurrent points on the main diagonal line are excluded when calculating the RQA measures, considering that each value within the sequence is recurrent with itself (Jenkins et al., 2020; Wallot, 2017).

Figure 1. The illustration of a recurrence plot



Note. The behavioral sequence is plotted on both the *x* and *y*-axis. The black dots and circles are placed in positions where the same behavior within the sequence reoccurs. The black circles form the main diagonal line, and the recurrence plot is symmetrical about its main diagonal line.

Researchers have successfully applied the RQA to investigate a wide range of social, physiological, psychological, and behavioral phenomena, such as process dynamics in organizations (Meinecke et al., 2020), heart rate variability (Marwan et al., 2002), cognition (Leonardi, 2012), emotion (Jenkins et al., 2020), and human behaviors (Fleuchaus et al., 2020; Wallot, 2017). For example, Jenkins et al. (2020) applied the RQA to analyze the temporal dynamics of affect, in particular, the degree of affect predictability, using the %*REC* and %*DET* measures as two crucial indicators. Fleuchaus et al. (2020) considered the %*LAM* measure as an indicator of behavioral stability, which varies from purely random to completely predictable. Specifically, Fleuchaus et al. (2020) used the %*LAM* measure to index the persistence of mistaken beliefs as students learned motor skills in science education.

Another representative example is the investigation of the motor-cognitive processes during a writing task, in which students were asked to copy-type a text (Wallot & Grabowski, 2019). When running RQA on students' keystroke logging data, all of the four RQA measures (i.e., %*REC*, %*DET*, *ADL*, and the maximum diagonal line length) suggested that copy-typing behaviors of a comprehensive text were more structured compared to that of an incomprehensible text. However, to the best of our knowledge, no research has used RQA to study students' self-regulated learning behaviors, especially in the context of clinical reasoning.

2.4. The current study

The present study situates the examination of the temporal structures and patterns of SRL behaviors in the context of clinical reasoning. Clinical reasoning is a complex thinking and decision-making process, in which medical practitioners extract meaningful information from patients' files, generate diagnostic hypotheses, order medical lab tests to confirm/disconfirm hypotheses, and finally propose a diagnostic solution (Eva, 2005). In clinical reasoning, medical students and practitioners are faced with a constant stream of decisions, which require them to effectively plan, monitor, control, and reflect on their behaviors. That is, SRL is an essential component for developing clinical competence.

While the research on medical students' SRL behaviors is emerging (Artino et al., 2011; Lajoie et al., 2021; Zheng et al., 2021), few studies have examined students' SRL behaviors with trace data that are collected in a specific clinical reasoning task. Contemporary research highlights the use of trace data since they capture the variations of students' SRL behaviors at a precise level of detail and consequently can afford a fine-grained level of analysis (Greene et al., 2019). What is more, the research on the temporal structures and patterns of SRL behaviors is still nascent in the context of clinical reasoning. This study aims to fulfill these gaps.

In particular, this study addresses the following two research questions: (1) Do high performers differ from low performers regarding the temporal structures of their SRL behaviors in clinical problem-solving? (2) Are there differences in the SRL behavioral patterns between high and low performers? Regarding the first research question, we cannot propose specific directional hypotheses since this study is one of the first to explore the RQA method in studying SRL behaviors, especially in the clinical reasoning context. But we assume that there will be significant differences between high and low performers in the temporal structures of SRL behaviors, i.e., the RQA measures of *%REC*, *%DET*, *ADL*, *%LAM*, and *TT*. For the second research question, we hypothesize that the SRL behavioral patterns of high performers demonstrate a better representation of general clinical reasoning procedures and are more closely aligned with the claims of SRL theories (i.e., the cyclical nature of SRL) than low performers.

3. Methods

3.1. Participants, learning environment, and task

In this study, the participants comprised of 75 medical students from a top North American university. There were 28 males (37.3%) and 47 females (62.7%), with an average age of 24.0 (*SD* = 3.17). Students were tasked to diagnose a virtual patient (VP) in BioWorld (Lajoie, 2020), which is an intelligent tutoring system that provides medical students with a safe practice environment for clinical reasoning. The main interface of the BioWorld platform is shown in Figure 2.

In BioWorld, students begin a task by reading a patient description, which contains information about the patient's profile and symptoms. Meanwhile, students collect evidence items from the patient description and store them in the *Evidence Table* for future reference. In doing so, students develop an initial understanding of the VP. They recognize what and how much information is important for the diagnosis since the *Evidence Table* functions as a metacognitive tool to help students monitor the process. Afterward, students can propose one or more diagnostic hypotheses. They confirm or disconfirm the proposed hypotheses by ordering medical lab tests, such as urine tests, toxicology tests, and blood sugar tests. Students can also search an online library with the platform if they need to acquire more information about a disease. Then, students link the collected evidence items and lab test results with corresponding diagnostic hypotheses. After submitting a final hypothesis, students check the relevance of evidence items and lab tests by categorizing them as either support, against, or neutral to the hypothesis. Students are also required to rank evidence items and lab tests based on their importance to the hypothesis. Finally, students write a case summary by reflecting on how they come up with the diagnosis.

The typical behaviors of clinical reasoning and how they relate to SRL phases were shown in Table 2 (Li et al., 2018). In particular, the behavior of collecting evidence items falls in the forethought phase, whereby students get familiar with the task conditions and develop an understanding of the patient problems. Students take actions to accomplish the task in the performance phase, which consists of three types of behaviors, i.e., raising/managing hypotheses, adding tests, and searching library. The self-reflection phase comprises four types of behaviors, namely, linking evidence/results, checking evidence/results, prioritizing evidence/results, and summarization for final diagnosis. These four types of behaviors involve extensive metacognitive activities as students evaluate the collected information, make adaptations and decisions, and reflect on their performance.

At the micro-level, the variations of SRL process are represented by the temporal changes in the eight types of clinical reasoning behaviors, which in consequence, trigger the changes in the three phases of SRL simultaneously at the macro-level. For instance, the transition from the behavior of collecting evidence items to adding tests occurs at the micro-level. Meanwhile, the behavioral transition indicates that students move from the forethought phase to the performance phase.

Figure 2. The main interface of the BioWorld environment

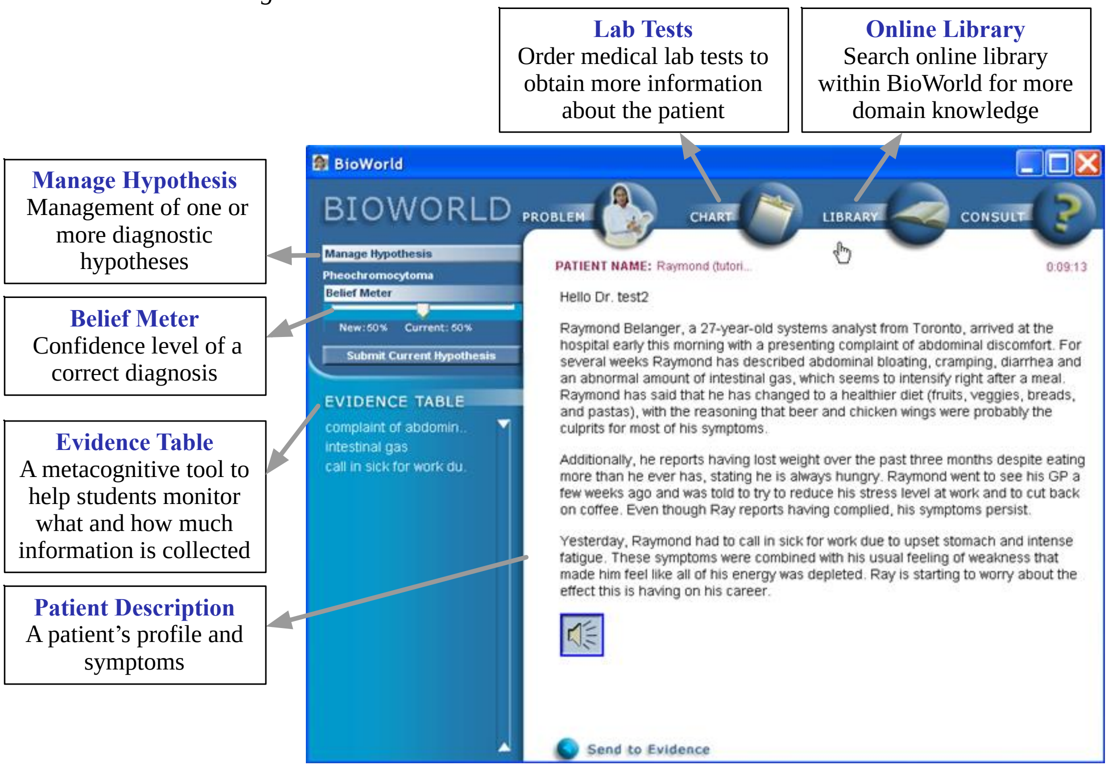


Table 2. The coding scheme for analyzing SRL behaviors of clinical reasoning

SRL phases	Clinical behaviors	Code	Description
Forethought	Collecting evidence items	CO	Collecting evidence items from the patient description by recalling one's prior knowledge pertaining to the symptoms
Performance	Raising/Managing hypotheses	RA	Outlining a single or multiple diagnostic hypothesis based on the collected evidence
	Adding tests	AD	Conducting medical lab tests
	Searching library	SE	Searching for information in the library for additional explanations
Self-Reflection	Linking evidence/results	LI	Linking evidence items and test results with corresponding diagnostic hypotheses
	Categorizing evidence/results	CA	Checking the relevance of evidence items and lab test results towards a specific hypothesis (i.e., whether the evidence/tests in support, against, or neutral of one hypothesis)
	Prioritizing evidence/results	PR	Ranking evidence items and lab test results according to their importance to a hypothesis
	Summarization for final diagnosis	SU	Making the final diagnosis by writing a summarization

The VP case used in this study was created by a panel of experts including medical professionals and learning scientists. The correct diagnosis for the VP case was pheochromocytoma, which is a rare, usually noncancerous tumor that develops in an adrenal gland. Prior to the study, a medical expert validated the VP case to ensure that it provided appropriate practice for the participants.

3.2. Procedure

Prior to the study, we had obtained research ethics approval from the institution's Research Ethics Board (REB) office. Moreover, we obtained the students' written consent to participate in this study. They all mentioned that they felt comfortable diagnosing virtual patients in BioWorld. Furthermore, they were informed that they had the right to withdraw at any time they wanted during the process of clinical problem-solving.

A training session was provided to help medical students get familiar with the BioWorld environment. In particular, the training session started with a researcher-guided introduction of the BioWorld system and how to use its various features to help them reach a final diagnosis. Afterward, the participants were asked to complete the clinical reasoning task independently. A number of research assistants were present to solve operational or technical issues; however, they were not allowed to give hints about the disease. The participants spent approximately 40 minutes on average to finish the diagnosis during regular school hours. It is worth mentioning that all operational behaviors (e.g., order lab test) and corresponding timestamps for each participant were automatically recorded in the log files of the BioWorld system.

3.3. Data processing and analysis

We first classified students as either high or low performers based on their diagnostic performance. Specifically, we considered students who correctly diagnosed the VP as high performers. The rest of the students were viewed as low performers since they failed to provide a correct diagnosis. In particular, there were 42 high performers and 33 low performers.

To address our first research question, we used the R package of "crqa" to perform the RQA on students' problem-solving behaviors separately for high and low performers (Coco & Dale, 2014). We then compared the differences in RQA measures (i.e., %*REC*, %*DET*, *ADL*, %*LAM*, and *TT*) between these two performance groups using inferential statistics. To address our second research question, we performed lag sequential analysis (Bakeman et al., 2009; Bakeman & Quera, 1995) to uncover the sequential behavioral patterns of both high and low performers. Specifically, we conducted the analysis using the GSEQ (Generalized Sequential Querier) program (Bakeman & Quera, 1995).

4. Results

4.1. Do high performers differ from low performers regarding the temporal structures of their SRL behaviors in clinical problem-solving?

The descriptive statistics of SRL behaviors were shown in Table 3. We compared the differences in the temporal structures of SRL behaviors (i.e., RQA measures) between high and low performers. As shown in Table 4, we found that there was no significant difference in the overall level of the repetitiveness of SRL behaviors between the two performance groups, $t(73) = .68$, $p = .499$. However, the %*DET* value for low performers ($M = 78.18$) was significantly lower than higher performers ($M = 86.37$), $t(73) = -3.35$, $p = .001$. This result indicated that low performers had a significantly higher ratio of single, isolated recurrent behaviors to all recurrent behaviors than high performers. The recurrent behaviors of high performers were more likely to be part of a behavioral sequence. The effect size ($d = .76$) was found to exceed Cohen's (1988) convention for a medium effect ($d = .50$).

Table 3. The descriptive statistics of SRL behaviors

	High performers				Low performers			
	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*
CO	8	34	15.14	3.79	10	34	14.91	5.25
RA	4	53	15.90	9.98	4	42	17.18	8.50
AD	4	48	18.95	11.03	0	57	15.85	11.94
SE	0	79	6.55	13.13	0	71	10.33	14.14
LI	0	69	17.43	16.74	0	68	14.00	14.52
CA	8	52	21.36	10.33	0	83	16.30	13.54
PR	3	102	30.69	27.49	0	267	32.12	50.89

Regarding the average length of the repeating behavioral patterns, there was no significant difference between high and low performers. There was also no significant difference in the average length of the repeating sequences of the same behavior between the two performance groups. Nevertheless, low performers showed a significantly lower ratio of repeating sequences of the same behavior to all recurrent behaviors than high performers, since the %*LAM* value for low performers (M = 86.23) was significantly lower than higher performers (M = 90.35), $t(73) = -2.26$, $p < .05$. This result suggested that high performers were more likely to perform a behavior repeatedly (e.g., order lab test) before moving on to the other behaviors than low performers. The effect size for the difference was medium, with Cohen's d = .52 (Cohen, 1988).

Table 4. Group differences in the temporal structures of SRL behaviors

	Group	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
%*REC*	Low	13.26	3.84	.68	73	.499	.15
	High	12.75	2.67				
%*DET*	Low	78.18	12.30	-3.35	73	.001**	.76
	High	86.37	8.85				
ADL	Low	5.74	3.91	-1.65	73	.104	.38
	High	7.05	2.94				
%*LAM*	Low	86.23	8.43	-2.26	73	.027*	.52
	High	90.35	7.35				
TT	Low	7.90	7.12	-1.57	73	.121	.36
	High	10.15	5.28				

Note. $^{*}p < .05$, $^{**}p < .01$. The definitions and meanings of the variables were shown in Table 1.

4.2. Are there differences in the SRL behavioral patterns between high and low performers?

As aforementioned, we performed behavioral sequential analyses for both high and low performers using the GSEQ program (Bakeman & Quera, 1995). Table 5 and Table 6 showed the sequential transition matrix of SRL behaviors of low performers and high performers, respectively. In the sequential transition matrix, the row means a starting behavior, whereas the column means a subsequent behavior. The values in the matrix are Z-scores. A Z-score greater than 1.96 indicates that the transition between two behaviors is statistically significant (i.e., $p < .05$) (Bakeman & Quera, 2011). Accordingly, a Z-score that is larger than 2.58 and 3.20 would guarantee significance levels of .01 and .001, respectively. For instance, the behavioral sequence of 'collecting evidence items → collecting evidence items', as shown in Table 5, was statistically significant, given that the Z-score = 53.44 > 3.20.

In fact, the results in Table 5 and Table 6 suggested that the sequential transitions between the same type of SRL behaviors were all statistically significant except the behavior of "summarization for final diagnosis." Moreover, the sequential transition from the behavior of "Prioritizing evidence/result" to "summarization for final diagnosis" was significant. Regarding the sequential transition patterns of SRL behaviors, there was no difference between low and high performers. The two performance groups both conducted a behavior repeatedly during the problem-solving process. However, high performers had a larger Z-score for each of the significant sequential transitions than low performers.

Table 5. The sequential transition matrix of SRL behaviors of low performers

Z	CO	RA	AD	SE	LI	CA	PR	SU
CO	**53.44**	-6.62	-6.14	-5.89	-7.94	-9.83	-12.39	-2.22
RA	-6.52	**40.62**	-5.99	-5.22	-2.55	-6.54	-13.46	-2.41
AD	-9.25	-1.81	**44.93**	.18	-8.85	-10.17	-12.82	-2.30
SE	-5.89	-4.91	1.16	**44.04**	-6.96	-8.00	-10.08	-1.81
LI	-8.02	-3.12	-8.71	-5.92	**51.32**	-9.48	-11.95	-2.14
CA	-9.44	-10.69	-10.17	-8.00	-9.48	**56.75**	-9.46	-2.34
PR	-11.91	-13.47	-12.82	-10.08	-11.95	-13.09	**57.84**	**10.92**
SU	.00	.00	.00	.00	.00	.00	.00	.00

Note. Z = Z-score, CO = Collecting evidence items, RA = Raising/Managing hypotheses, AD = Adding tests, SE = Searching library, LI = Linking evidence/results, CA = Categorizing evidence/results, PR = Prioritizing evidence/results, SU = Summarization for final diagnosis. Only significant sequential transitions were highlighted in bold.

Table 6. The sequential transition matrix of SRL behaviors of high performers

Z	CO	RA	AD	SE	LI	CA	PR	SU
CO	**64.30**	-5.61	-6.34	-5.23	-10.35	-11.99	-14.99	-2.37
RA	-6.57	**50.12**	-7.34	-4.23	-3.15	-7.89	-15.66	-2.47
AD	.00	-3.38	**58.72**	1.63	-11.63	-13.91	-17.39	-2.75
SE	-4.56	-2.54	.60	**51.70**	-6.48	-7.74	-9.68	-1.53
LI	-9.05	-4.70	-11.52	-5.59	**63.83**	-13.24	-16.56	-2.61
CA	-11.53	-12.51	-13.91	-7.74	-13.24	**68.50**	-15.15	-2.96
PR	-14.42	-15.64	-17.39	-9.68	-16.56	-18.73	**69.44**	**11.46**
SU	.00	.00	.00	.00	.00	.00	.00	.00

Note. Only significant sequential transitions were highlighted in bold.

In addition, we examined the sequential transitions between different SRL behavioral states by viewing temporally connected behaviors of the same type (e.g., CO → CO → CO) as a behavioral state, i.e., the state of CO. As an illustration, the behavioral sequence of "CO → CO → CO → RA → AD → AD" would be transformed as "CO → RA → AD" for further analysis. In doing so, we may develop a deep understanding of students' thinking and reasoning activities as they are reflected by different behavioral states. Additionally, new insights may be obtained by examining the sequential transitions of different behavioral states, given that the sequential transitions between the same type of SRL behaviors were all found to be statistically significant.

The sequential transition matrices of SRL behavioral states of low performers and high performers were shown in Table 7 and Table 8, respectively. Based on the sequential transition matrices, we made the transition diagrams of SRL behavioral states for the two performance groups (see Figure 3 and Figure 4 the diagrams of low and high performers, respectively).

Table 7. The sequential transition matrix of SRL behavioral states of low performers

Z	CO	RA	AD	SE	LI	CA	PR	SU
CO	0	.21	3.00^{**}	-.08	-.11	-2.03	-2.03	-1.98
RA	3.13^{**}	0	-3.35	-3.46	6.87^{***}	7.23^{***}	-4.17	-4.07
AD	-3.34	2.51^{*}	0	7.37^{***}	-3.72	-3.09	-3.09	-3.02
SE	1.18	-3.06	8.96^{***}	0	-2.94	-2.44	-2.44	-2.39
LI	.04	5.83^{***}	-3.41	-.76	0	-2.01	-2.01	-1.96
CA	-1.45	-4.18	-3.09	-2.44	-2.01	0	25.51^{***}	-1.31
PR	-1.45	-4.18	-3.09	-2.44	-2.01	-1.34	0	26.20^{***}
SU	0	0	0	0	0	0	0	0

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 1. The sequential transition matrix of SRL behavioral states of high performers

Z	CO	RA	AD	SE	LI	CA	PR	SU
CO	0	.94	6.13^{***}	-1.6	-2.69	-2.43	-2.43	-2.36
RA	2.18^{*}	0	-3.48	-4.12	8.09^{***}	8.02^{***}	-4.81	-4.67
AD	-3.16	2.21^{*}	0	9.62^{***}	-3.23	-3.49	-3.49	-3.39
SE	1.05	-2.01	7.90^{***}	0	-2.66	-2.4	-2.4	-2.33
LI	2.10^{*}	5.32^{***}	-3.23	-0.77	0	-2.38	-2.38	-2.31
CA	-1.7	-4.8	-3.49	-2.4	-2.38	0	26.43^{***}	-1.63
PR	-1.7	-4.8	-3.49	-2.4	-2.38	-1.67	0	27.30^{***}
SU	0	0	0	0	0	0	0	0

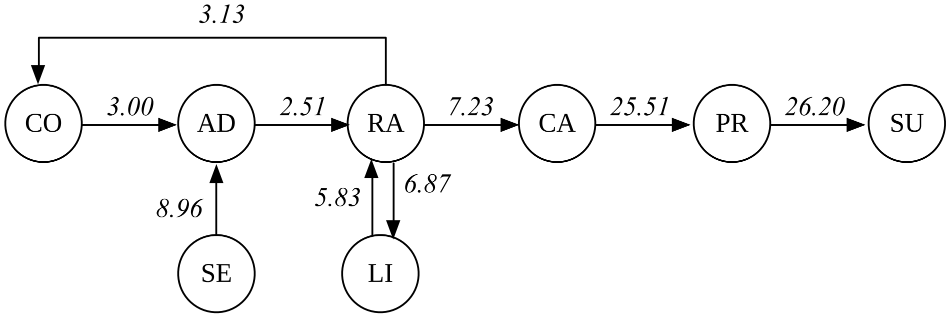
Note. $^{*}p < .05$, $^{***}p < .001$.

As shown in

and Figure, low and high performers demonstrated similar behavioral patterns in general. They both demonstrated a behavioral pattern of "CO (Collecting evidence items) → AD (Adding tests) → RA (Raising/Managing hypotheses) → CA (Categorizing evidence/results) → PR (Prioritizing evidence/results) → SU (Summarization for final diagnosis)." Moreover, the sequential transition from the behavior of "Raising/Managing hypotheses" to "Collecting evidence items" was significant for both low- and high-performing groups. In addition, both low and high performers conducted the behavior of "Raising/Managing hypotheses" following the behavior of "Linking evidence/results," and vice versa. Nevertheless, there was a reciprocal sequential transition between the behaviors of "Searching library" and "Adding tests" for high performers, while the transition between the two behaviors was unidirectional for low performers. It was also noticeable that the behavior of "Linking evidence/results" significantly stimulated the occurrence of the

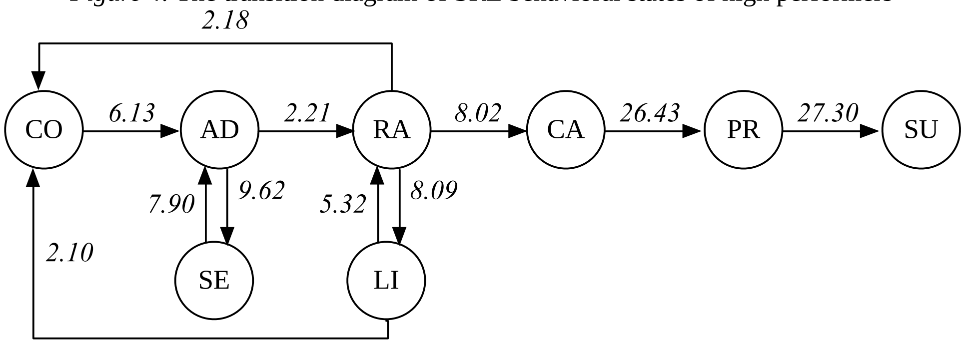
"Collecting evidence items" behavior for high performers, whereas low performers showed no such a behavioral pattern.

Figure 3. The transition diagram of SRL behavioral states of low performers



Note. Only significant sequential transitions of SRL behavioral states were displayed in the Figure. The numbers above the directional lines were Z-scores, with a larger value indicating a stronger relationship between two SRL behavioral states.

Figure 4. The transition diagram of SRL behavioral states of high performers



5. Discussion

In this study, we found that low performers had more single, isolated recurrent behaviors in problem-solving, whereas the recurrent behaviors of high performers were more likely to be part of a behavioral sequence. This finding is aligned with the research on the development of professional expertise (Lajoie, 2009). High performers may use their mental models, which can be developed from either experience or instruction, to drive the selection of problem-solving behaviors. In other words, high performers may have developed internal representations of the task condition and see relevant concept relationships (He et al., 2021). Consequently, they follow specific implicit procedures to perform the task. It is also quite possible that high performers are self-regulated learners who can effectively monitor and control their problem-solving processes. High performers are aware of the next desirable behavior based on the outcome of prior behavior, thus yielding more meaningful behavioral sequences than low performers.

While these explanations are grounded in solid theoretical frameworks and in common sense, researchers should not draw a conclusion that this finding is applicable to the research contexts other than clinical reasoning. In fact, we argued that this finding might be confined to the contexts or disciplines that are governed by facts, principles, rules, and scientific reasoning. Take this study as an example, high performers demonstrated more behavioral patterns, because the procedures of clinical reasoning are well-established, and physicians are aware of those procedures. In the learning or problem-solving contexts that require creativity and innovation, the lack of regularity in students' learning behaviors, however, may be a crucial indicator of high performance, since students need to challenge their conventional thinking for breakthrough success (Koopmans, 2020).

This study found that high performers were more likely than low performers to perform a behavior repeatedly before moving on to other behaviors. This finding can be explained by the fact that clinical reasoning in medicine traditionally values the ultimate goal of providing accurate diagnoses of disease (Li et al., 2020; Wass et al., 2001). High performers may purposefully conduct a behavior repeatedly to eliminate any sense of uncertainty in the process of clinical problem-solving. For example, they may conduct multiple diagnostic tests needed to narrow down a diagnosis. Additionally, each step of clinical reasoning has its own challenges. For instance, medical students may have difficulties in identifying cues, developing an understanding of patient problems, generating diagnostic hypotheses, prioritizing evidence items, and finalizing a decision (Audétat et al., 2017; Gonzalez et al., 2021). Therefore, it is reasonable to infer that high performers were more behaviorally engaged than low performers at each step of clinical reasoning to address ambiguities. Clearly, more research is needed to uncover the underlying mechanisms for this finding by fusing both objective (log files) and subjective data (direct input from the participants such as self-reports, think-aloud, or interviews).

Interestingly, both low and high performers demonstrated clear and logical transitions between and among SRL behavioral states revealing similar behavioral patterns in clinical problem-solving. In particular, the behavioral state transitions of both low and high performers were in compliance with a loosely sequential process of SRL, i.e., forethought, performance, and self-reflection (Schunk & Greene, 2017; Zimmerman, 2000). This result was consistent with the research of Greene et al. (2019), who also found evidence of temporality in SRL. Nevertheless, the sequential transitions tell a more in-depth story about the relationship between SRL and performance.

Findings from the sequential transition analyses revealed that the probability of the sequential transition from the forethought phase to the performance phase (i.e., collecting evidence items → adding tests) was higher for high performers than that of low performers. Moreover, high performers were more likely to proceed to the self-reflection phase (i.e., linking, categorizing, and prioritizing evidence/results) from the performance phase (i.e., raising/managing hypotheses) than low performers. High performers were also unique in that their behavioral state transitions were cyclically sustained, as suggested by a behavioral state transition from "linking evidence/results" (i.e., self-reflection phase) to "collecting evidence items" (i.e., forethought phase). The feedback loop may help high performers adaptively adjust their SRL behaviors over cycles, allowing them to navigate their decision-making processes towards a correct diagnosis.

In addition, the sequential transition between the behaviors of "searching library" and "adding tests" was bidirectional for high performers, while the transition between the two behaviors was unidirectional for low performers, i.e., searching library → adding tests. This finding is partially in line with our previous studies, in which we found that learners would commence by ordering a lab test prior to a library search (Li et al., 2020), and high performers spent significantly more time ordering lab tests than low performers before developing their first diagnostic hypothesis (Li et al., 2020). In essence, the two behaviors both fell into the performance phase of SRL. When students were unfamiliar with a specific disease or diagnostic tests, they searched the library to gain more information. They ordered lab tests to confirm or rule out a specific hypothesis regarding which disease was present. Both high and low performers consulted the online library to guide the selection of lab tests. However, high performers tended to search library to clarify the meaning of the results of the lab test after they collected the test. As a result, high performers were able to take corrective action in a timely fashion when ordering the next lab test, as they gained additional information of a prior test from searching library.

Moreover, according to Zimmerman (2000), attentional control is a crucial strategy used intensively by expert performers in the performance phase of SRL. In the same vein, high performers knew how to concentrate in the performance phase of clinical reasoning by ignoring distractions and by focusing their attention on those two behaviors exclusively, which could also explain the reciprocal relationship between the two behaviors.

In summary, the findings from this study uphold the theoretical assumption that students' SRL behaviors occur in a loosely sequenced and temporal order (Bernacki, 2017; Broadbent et al., 2020; Winne, 2019). Moreover, the results of the sequential pattern analyses revealed that only high performers repeatedly progressed through the three SRL phases (i.e., forethought, performance, and self-reflection) in a cyclical and iterative fashion, which may help explain the performance difference with low performers. A unique theoretical contribution of this study is that high performers were found to have a higher transition probability across the three SRL phases than low performers. In this regard, this study informs future research on the theoretical advancements of SRL by examining how the three phases of SRL and even the behaviors within each phase are interconnected.

Furthermore, this study provided significant methodological insights regarding the quantification of the temporal structure of SRL behaviors. Specifically, this study is one of the first to demonstrate how RQA can be used to quantify the temporal structure of SRL behaviors in the context of clinical reasoning. Along with the RQA measures that describe the overall characteristics of students' SRL behaviors (e.g., repetitiveness and predictability), we looked into the sequential patterns of those SRL behaviors using sequential analysis (Bakeman et al., 2009). The use of both RQA and sequential analysis provided researchers a complete picture of students' SRL behaviors, whereby new understandings of students' performance differences can be obtained.

In addition, the present research has practical implications. Specifically, findings from this study can inform the design of learning analytics dashboards, which provide educators and students with the opportunities for awareness, reflection, sensemaking, and behavioral change. For instance, educators can develop an understanding of students' potential performance from their behavioral characteristics so that they can adjust their instructional strategies accordingly. Moreover, educators need to carefully design their course activities to facilitate the acquisition of not only domain-specific knowledge but also SRL skills among students. This study also informs the design of early warning systems (e.g., automatic detection of at-risk students from analyzing their behavioral patterns), whereby immediate support can be provided to help all students succeed.

6. Conclusion

In this study, we examined the temporal structures and patterns of SRL behaviors, as 75 medical students solved a clinical reasoning task within an intelligent tutoring system. We found that the recurrent behaviors of high performers were more structured and predictive than low performers. As revealed by the sequential pattern analysis, high performers also demonstrated a higher transition probability across the three phases of SRL than low performers. Moreover, high performers were unique in that their behavioral state transitions were cyclically sustained. In addition to its methodological insights that could inform future research significantly, this study provided researchers with new evidence to support the theoretical assumptions of SRL in the context of clinical reasoning with the population of medical students.

Nevertheless, this study is not without limitations. First, we did not explicitly assess students' prior knowledge and skills, although we had confirmed with a medical expert to ensure the appropriateness of the task for the participants. Second, we had to make inferences about students' decision-making process from log files, as log files do not directly measure cognitive and metacognitive activities (Paans et al., 2019). Consequently, we identified the characteristics and patterns of students' SRL behaviors; however, our explanations about the research findings called for further studies. Third, we compared the behavioral patterns of high and low performers; however, we could not tell which differences in the behavioral patterns were statistically significant between those two groups. Lastly, the participants were all from the same university, which may affect the generalizability of our research findings.

Despite these limitations, this research opens new directions for future research. For one, it would be fruitful to examine the relationships between the temporal structures of SRL behaviors (i.e., RQA measures) and other psychological or contextual factors, such as students' personality, motivation, and emotion. Task difficulty is another crucial fact that may affect the characteristics and patterns of SRL behaviors. Therefore, an important future direction is to examine the influences of task difficulty on the temporal structures and patterns of SRL behaviors. It is also promising to study SRL behaviors in other learning or problem-solving contexts using RQA and sequential behavioral analysis simultaneously in one study.

References

Artino, A. R., Hemmer, P. A., & Durning, S. J. (2011). Using self-regulated learning theory to understand the beliefs, emotions, and behaviors of struggling medical students. *Academic Medicine*, *86*(10), S35–S38. https://doi.org/10.1097/ACM.0b013e31822a603d

Audétat, M.-C., Laurin, S., Dory, V., Charlin, B., & Nendaz, M. R. (2017). Diagnosis and management of clinical reasoning difficulties: part II. Clinical reasoning difficulties: management and remediation strategies. *Medical Teacher*, *39*(8), 797–801. https://doi.org/10.1080/0142159X.2017.1331034

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self-and socially-regulated learning. *Metacognition and Learning*, *9*(2), 217–228. https://doi.org/10.1007/s11409-014-9123-1

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. Cambridge University Press.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.

Bakeman, R., Quera, V., & Gnisci, A. (2009). Observer agreement for timed-event sequential data: A Comparison of time-based and event-based algorithms. *Behavior Research Methods*, *41*(1), 137–147. https://doi.org/10.3758/BRM.41.1.137

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185. https://doi.org/10.1007/s11409-013-9107-6

Bernacki, M. L. (2017). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 370–387). Routledge/Taylor & Francis Group.

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. Paul R, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (1st ed., pp. 417–450). US: Academic Press. https://doi.org/10.1016/B978-012109890-2/50042-1

Chen, C.-H., & Su, C.-Y. (2019). Using the BookRoll e-book system to promote self-regulated learning, self-efficacy and academic achievement for university students. *Journal of Educational Technology & Society*, *22*(4), 33–46.

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology*, *5*, 1–14. https://doi.org/10.3389/fpsyg.2014.00510

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Inc.

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education*, *39*(1), 98–106. https://doi.org/10.1111/j.1365-2929.2004.01972.x

Fleuchaus, E., Kloos, H., Kiefer, A. W., & Silva, P. L. (2020). Complexity in science learning: Measuring the underlying dynamics of persistent mistakes. *Journal of Experimental Education*, *88*(3), 448–469. https://doi.org/10.1080/00220973.2019.1660603

Gonzalez, L., Nielsen, A., & Lasater, K. (2021). Developing students' clinical reasoning skills: A Faculty guide. *Journal of Nursing Education*, *60*(9), 485–493. https://doi.org/10.3928/01484834-20210708-01

Gottman, J. M., Gottman, J. M., & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers*. Cambridge University Press.

Greene, J. A., & Azevedo, R. (2007). A Theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, *77*(3), 334–372. https://doi.org/10.3102/003465430303953

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., Demetriou, C., & Panter, A. T. (2019). Modeling temporal self-regulatory processing in a higher education biology course. *Learning and Instruction*, 101201. https://doi.org/10.1016/j.learninstruc.2019.04.002

He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, *166*, 104170. https://doi.org/10.1016/j.compedu.2021.104170

Jenkins, B. N., Hunter, J. F., Richardson, M. J., Conner, T. S., & Pressman, S. D. (2020). Affect variability and predictability: Using recurrence quantification analysis to better understand how the dynamics of affect relate to health. *Emotion*, *20*(3), 391–402. https://doi.org/10.1037/emo0000556

Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, *6*(1), 39–56. https://doi.org/10.1007/s11412-011-9109-9

Koopmans, M. (2020). Education is a complex dynamical system: Challenges for research. *Journal of Experimental Education*, *88*(3), 358–374. https://doi.org/10.1080/00220973.2019.1566199

Kuvalja, M., Verma, M., & Whitebread, D. (2014). Patterns of co-occurring non-verbal behaviour and self-directed speech; a comparison of three methodological approaches. *Metacognition and Learning*, *9*(2), 87–111. https://doi.org/10.1007/s11409-013-9106-7

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.), *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments* (pp. 61–83). Cambridge University Press.

Lajoie, S. P. (2020). Student modeling for individuals and groups: the BioWorld and HOWARD platforms. *International Journal of Artificial Intelligence in Education*, 1–16. https://doi.org/10.1007/s40593-020-00219-x

Lajoie, S. P., Li, S., & Zheng, J. (2021). The Functional roles of metacognitive judgement and emotion in predicting clinical reasoning performance with a computer simulated environment simulated environment. *Interactive Learning Environments*, 1–12. https://doi.org/10.1080/10494820.2021.1931347

Leonardi, G. (2012). The study of language and conversation with recurrence analysis methods. *Psychology of Language and Communication*, *16*(2), 165. https://doi.org/10.2478/v10057-012-0012-x

Li, S., & Lajoie, S. P. (2021). Cognitive engagement in self-regulated learning: An Integrative model. *European Journal of Psychology of Education*, 1–20. https://doi.org/10.1007/s10212-021-00565-x

Li, S., Zheng, J., & Lajoie, S. P. (2020). Efficient clinical reasoning: Knowing when to start and when to stop. *Education in the Health Professions*, *3*(1), 1–7.

Li, S., Zheng, J., Poitras, E., & Lajoie, S. (2018). The Allocation of time matters to students' performance in clinical reasoning. In R. Nkambou, R. Azevedo, & J. Vassileva (Eds.), *Lecture notes in computer sciences* (pp. 110–119). Springer International Publishing AG, part of Springer Nature. https://doi.org/10.1007/978-3-319-91464-0_11

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., & Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, *66*(2), 26702. https://doi.org/10.1103/PhysRevE.66.026702

Meinecke, A. L., Handke, L., Mueller-Frommeyer, L. C., & Kauffeld, S. (2020). Capturing non-linear temporally embedded processes in organizations using recurrence quantification analysis. *European Journal of Work and Organizational Psychology*, *29*(4), 483–500. https://doi.org/10.1080/1359432X.2019.1658624

Michailidis, N., Kapravelos, E., & Tsiatsos, T. (2018). Interaction analysis for supporting students' self-regulation during blog-based CSCL activities. *Journal of Educational Technology & Society*, *21*(1), 37–47.

Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2019). Temporal variation in children's self-regulated hypermedia learning. *Computers in Human Behavior*, *96*, 246–258. https://doi.org/10.1016/j.chb.2018.04.002

Panadero, E. (2017). A Review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Pintrich, P. R. (2004). A Conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*(4), 1321–1331. https://doi.org/10.1016/j.chb.2012.02.016

Schunk, D. H., & Greene, J. A. (2017). Historical, contemporary, and future perspectives on self-regulated learning and performance. In P. A. Alexander, D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 1–15). Routledge. https://doi.org/10.4324/9781315697048-1

Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A Tutorial in R. *Discourse Processes*, *54*(5–6), 382–405. https://doi.org/10.1080/0163853X.2017.1297921

Wallot, S., & Grabowski, J. (2019). A Tutorial introduction to recurrence quantification analysis (RQA) for keystroke logging data. In E. Lindgren & K. Sullivan (Eds.), *Observing Writing* (pp. 163–189). Brill. https://doi.org/10.1163/9789004392526

Wass, V., Van Der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *Lancet*, *357*(9260), 945–949. https://doi.org/10.1016/S0140-6736(00)04221-5

Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In P. A. Alexander, D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 36–48). Routledge. https://doi.org/10.4324/9781315697048-3

Winne, P. H. (2019). Paradigmatic dimensions of instrumentation and analytic methods in research on self-regulated learning. *Computers in Human Behavior*, *96*, 285–289. https://doi.org/10.1016/j.chb.2019.03.026

Zheng, J., Li, S., & Lajoie, S. P. (2021). Diagnosing virtual patients in a technology-rich learning environment: A Sequential mining of students' efficiency and behavioral patterns. *Education and Information Technologies*. https://doi.org/10.1007/s10639-021-10772-0

Zimmerman, B. J. (2000). Attaining self-regulation: A Social cognitive perspective. In M. Boekaerts, P. Paul R, & M. Zeidner (Eds.), *Handbook of self-regulation* (1st ed., pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Seo, M.-H., & So, H.-J. (2022). Developing a Gesture-based AR Exhibit: Differently-Guided Experiences for Complex Conceptual Learning in Science. *Educational Technology & Society*, *25* (4), 15-28.

Developing a Gesture-based AR Exhibit: Differently-Guided Experiences for Complex Conceptual Learning in Science

Min-Hwi Seo and Hyo-Jeong So*
Department of Educational Technology, Ewha Womans University, Seoul, South Korea // minhwiseo@ewha.ac.kr // hyojeongso@ewha.ac.kr
*Corresponding author

(Submitted May 4, 2021; Revised November 30, 2021; Accepted December 27, 2021)

ABSTRACT: The purpose of this research was to design and evaluate the efficacy of a gesture-based exhibit with augmented reality (AR) for understanding complex scientific concepts. In particular, this study focuses on the effect of differently guided conditions in a gesture-based AR. We first present the design and development of a gesture-based AR exhibit about the conductor resistance phenomenon. An experiment was conducted to examine the effect of guided and unguided experiences on complex conceptual learning. In the experiment, 40 participants between 15 and 17 years-old were randomly assigned to either the guided (visual and docent explanation) or unguided condition. Their understanding of complex concepts was measured through the pre-test and post-test. The results indicate that while the participants increased cognitive understanding after experiencing the gesture-based AR exhibit, there was no significant difference between the two conditions. This may imply that the provision of extra guidance does not necessarily lead to better conceptual learning. In conclusion, this study provides some implications concerning the design of new types of immersive exhibits in museum contexts.

Keywords: Augmented reality, Informal learning, Science museum, Conceptual learning

1. Introduction

One of the main goals of science museums is to help visitors understand scientific phenomena and principles. Recently, science museums have been transformed by integrating emerging technologies in exhibit design. Underlying this transformation is the shift from object-based design to visitor experience design (Matuk, 2016). Object-based design refers to the use of real objects for knowledge transmission. As contemporary museums are increasingly concerned about supporting constructivist learning goals, it became important to design visitor experiences to be participatory, interactive, and immersive. This supports visitors in constructing knowledge based on their own experiences, interpretation, and perspectives (Freeman et al., 2016; Matuk, 2016). Augmented reality (AR) that integrates the physical and digital worlds has been proposed as a relevant strategy for designing interactive and immersive visitor experiences.

This study investigates two fundamental factors related to the design of immersive visitor experiences: the use of gestures and the provision of guidance. These two factors are related to the cognitive load that users may encounter during immersive experiences, which is one of the most critical design challenges (Dunleavy, 2014; Yoon et al., 2013). First, we suggest that gesture interaction using the body as an input source can be an effective design strategy to reduce potential cognitive load and to allocate more cognitive resources to higher-level learning processes. Gestures as a new type of user interface enable more intuitive and natural interaction (Fang et al., 2007). In problem-solving situations, gesturing reduces demands on cognitive resources and permits the allocation of more resources to perform tasks (Goldin-Meadow et al., 2001). Furthermore, conceptual learning can be enhanced when there is a high congruency between gestures and concepts to be learned (Antle et al., 2008; Han & Black, 2011). Despite such potential, little is known about the effects of gesture-based systems on cognitive learning, especially in the context of immersive learning environments.

Second, there have been attempts to reduce cognitive load in AR-integrated systems with the provision of guidance (Matuk, 2016). Dunleavy (2014) summarizes design strategies in AR-based learning to minimize cognitive load as (a) creating a simplified experience structure and increasing complexity gradually, (b) providing scaffolds that guide learning processes explicitly to achieve desired goals, and (c) replacing text with audio and video narrations. Although these strategies may be effective, there are also concerns about "overformalization" of learning experiences in highly-scaffolded AR-based learning (Yoon et al., 2013). This issue is particularly important in informal learning settings such as science museums since flexible and voluntary participation is central to visitors' experiences. For instance, Yoon et al. (2013) found that as scaffolds increased, the level of informal participation behaviors decreased in the science museum.

The issue of guidance and scaffolding is relevant to the design of interactive AR exhibits since managing users' cognitive load is a critical design challenge (Baydas et al., 2015; Dunleavy, 2014; Matuk, 2016). Whereas the tension between guidance and learning has been extensively debated in the literature on instructional approaches (e.g., Hmelo-Silver et al., 2007; Kirschner et al., 2006; Kuhn, 2007; Tobias & Duffy, 2009), little is known about this issue in the context of exhibit design and AR-based learning. Furthermore, previous research indicates the inherent dilemma of interactive exhibits. Allen (2004) suggests the need to take a critical stance on the dilemma between interactive elements and learning, by questioning whether integrating various interactive elements in exhibit design promotes more audience participation and better learning experiences. It is critical to go beyond the simple approach of "more is better" and to deeply examine optimal conditions for interactivity and guidance.

With this backdrop, the purposes of this study are to develop a gesture-based AR exhibit about complex science concepts and to evaluate its effect in differently guided conditions. The research question examined is "is there any significant difference between guided and unguided experiences in terms of the effect on complex conceptual learning?" Based on the lack of empirical research on guidance conditions in AR-based learning, we explore whether learning outcomes differ between guided and unguided experiences. This paper first presents the design and development of the gesture-based AR exhibit on the topic "Current and Resistance" to help users learn about complex invisible concepts. Then, we present a quasi-experimental study that investigated the effect of the gesture-based AR exhibit under differently-guided conditions on complex conceptual learning. The guidance used in the experiment included a visual explanatory panel and verbal explanation by a docent, the latter being the most commonly used technique for guidance in museum settings. Based on the key research findings, we attempt to draw some implications concerning designing new types of immersive exhibits in museum contexts.

2. Theoretical backgrounds

2.1. Interactive exhibit and gestures

According to Gardner (1995), an interactive exhibit is one type of participatory exhibit that allows visitors to directly experience the content in a way they can understand and appreciate. Gardner proposes four types of participatory exhibits depending on the presentation style and function: (a) hands-on exhibits, (b) interactive exhibits that stimulate sensory organs, (c) eyes-on exhibits that visitors observe visually, and (d) exhibits made up of design panels and simple image panels. Museum exhibits are moving from eyes-on exhibits to interactive exhibits by experimenting with emerging technologies for visitor engagement (Goff et al., 2018). In particular, gesture-based computing can support various physical interactions among visitors (Matuk, 2016).

Embodied cognition is essential to the design of gesture-based exhibits. While the conventional concept of cognition emphasizes cognitive representations and information processing in the human brain, embodied cognition holds that cognition is connected to the environment through bodily gestures rather than being an abstract proposition in the brain (Wilson, 2002). Wilson and Foglia (2017) contend that personal perceptions are based on body movements, which have a significant impact on visual attention, concept, memory, understanding of others, and even moral perception. Previous research has reported that gesture-based learning is effective in acquiring higher-level concepts and correcting misconceptions in science learning (Han & Black, 2011). For instance, Goldin-Meadow et al. (2001) reported that when learners were asked to memorize and explain a list of items, the performance level of learners who were allowed to use gestures was higher than that of learners who were restricted with the use of gestures.

The review of previous research on interactive exhibits and gesture-based exhibits reveals certain features that are likely to affect the success of visitor experiences. Here, we discuss three key features, which also informed the design of the gesture-based AR exhibit proposed in this study. First, an interactive design that promotes social interaction tends to promote more participation and engagement. For instance, Horn et al. (2012) analyzed family visitors using an interactive tabletop game in a natural history museum and found that gameplay elements significantly contributed to visitors' collaborative conversation, which subsequently influenced active prolonged engagement (APE) with the exhibit. Hinrichs and Carpendale (2011) also reported that the use of multi-touch gestures on an interactive table exhibit facilitated the emergence of social information exploration in the aquarium setting.

Second, it is important to provide natural mappings between gestures and concepts—meanings conveyed in the exhibit. The notion of "embodied metaphorical mappings" (Antle et al., 2008) is relevant for understanding the relationship between gestures and learning. Antle et al. (2008) introduce embodied metaphorical mappings to indicate that certain physical movements have metaphorical meanings that are cognitively mapped. For example, if an exhibit requires a movement to manipulate speed by manipulating tempo, it implicitly conveys the metaphorical meaning, "When the tempo gets faster, the speed gets faster. When the tempo slows, the speed slows down." As another example, many of the hands-on exhibits use power as input. These exhibits convey a built-in metaphorical meaning that an observer's "exertion" action consumes or accumulates more power or data (Lyons et al., 2012).

Third, interactive exhibits with manipulative experiences do not necessarily lead to enhanced immersion and interest of visitors, and sometimes can cause confusion and misconceptions. For example, the "Hot and Cold Coils" displayed at the Exploratorium provides visitors with learning opportunities through direct touching and manipulating of the exhibit. Despite such interactive experiences, Gutwill and Allen (2012) found that the length of time that visitors stay for the experience is rather short and that some visitors repeat meaningless actions. In the subsequent section, we further discuss this dilemma in interactive exhibits concerning the provision of guidance and visitor experiences, which is the central topic of our investigation.

2.2. Guidance in visitor experiences

With the movement toward constructivist learning goals, there have been extensive discussions about guided learning versus unguided or minimally-guided learning (Hmelo-Silver et al., 2007; Kirschner et al., 2006; Kuhn, 2007; Tobias & Duffy, 2009). Kirschner et al. (2006) define minimally-guided learning as "an approach where learners, rather than being presented with essential information, must discover or construct essential information for themselves" (p. 1), which is the method emphasized in constructivist learning approaches such as inquiry learning (IL) and problem-based learning (PBL). On the other hand, direct instruction refers to "providing information that fully explains the concepts and procedures that students are required to learn" (p. 1). Kirschner et al. (2006) argue that approaches under minimally-guided instruction are less effective than guided instruction since minimally-guided instruction poses heavy demands on working memory, especially for novice learners who lack relevant prior knowledge. Several scholars such as Hmelo-Silver et al. (2007) and Kuhn (2007), however, disagree with that argument and contend that PBL and IL are not minimally-guided learning, but embed many forms of scaffolding (e.g., benchmark lesson and just-in-time support) to help learners understand necessary disciplinary knowledge and also better manage problem-solving processes.

At science museums, guidance is provided in various forms including docents, labels, commentary panels, and audio narrations. These guiding methods can be broadly classified into linguistic, visual, and auditory scaffolding. In general, the role of guidance in visitor experiences can be understood for two purposes: to reduce cognitive load and to promote better conceptual understanding. First, visitors' learning from exhibits is achieved through cognitive information processing. Visitors selectively perceive sensory information to interpret the meaning of exhibits. The perception and reaction of visitors can change depending on how their senses perceive the exhibit at a sensory memory stage (Moreno, 2004). However, cognitive load may occur when information processing is concentrated on certain senses (Sweller et al., 1998). Subsequently, this cognitive load increases visitors' fatigue, which makes them either give up or avoid devoting cognitive resources to exhibit experiences. When information is presented in multiple forms, cognitive load can be reduced by dispersing them in the linguistic and visual processing of the working memory (Mayer & Moreno, 2003; Sweller et al., 1998). For instance, Sun and You (2019) found that personalized museum guides aligned with learning styles (e.g., visualizer or verbalizer) reduce cognitive load and help visitors better remember information.

Second, visitors are provided with appropriate guidance or explanation of exhibits to improve their understanding of objects, natural phenomena, or scientific principles. We discuss two forms of guidance frequently used in exhibit design: explanatory panel and docents. An explanatory panel as linguistic and visual scaffolding is the most commonly-used guidance in exhibit design. Falk (1997) reported that explanatory panels about scientific concepts are effective in conceptual learning. Similarly, Hohenstein and Tran (2007) found that explanatory panels that pose inquiry questions stimulate the audience's open discourse. Some studies, however, question the efficacy of linguistic scaffolds. Allen and Gutwill (2004) suggest that labels in interactive exhibits are not always useful because it is difficult for visitors to clearly understand the meaning that exhibit labels convey. The efficacy of explanatory panels is mainly evaluated in terms of *attracting power* and *holding power*. Attracting power refers to how many visitors pay attention to the explanation panel, whereas holding power refers to reading time. Korn and Jones (2000) argue that an explanation panel dealing with in-depth scientific concepts rarely draws visitors' attention, implying low attracting power. However, once visitors are attracted to the panel, it can maintain the viewer's attention for a certain period, thereby increasing holding power.

In addition to explanatory panels, museum docents have been suggested as effective for promoting cognitive learning (Braund & Lelliott, 2017; King & Tran, 2017; Shaby et al., 2019). For instance, Hooper-Greenhill (1999) suggested that the cognitive and affective stimuli provided by docents promote more active visitor-exhibit interaction. Guided learning by museum docents, however, can lead to different effects in cognitive and affective areas. A classic study by Stronck (1983) compared the effects between the structured tour guided by a docent and the less-structured tour guided by a school teacher. The results indicate that a guided tour by docents is effective for cognitive learning while a less-structured tour is more effective for promoting students' positive attitudes.

Concerning the tension between free-choice learning and structured learning, Gutwill and Allen (2012) conducted an experiment where visitors were randomly assigned to four conditions where the degree of structure and collaboration differ. They found that learning gains were higher under structured and collaborative conditions than under spontaneous and individualized conditions. Some studies, however, have reported that desired learning outcomes can be achieved under minimally-guided situations. For instance, Yasar and Gurel (2016) contend that the learning effect and visitors' interest can be high when there is a tight coupling between visitor behaviors and learning content induced by the exhibit design. The research on Physics Education Technology (PhET) also suggests that when an interactive simulation is designed to make conceptual models used by experts visible, it can increase learners' understanding of complex concepts (Wieman et al., 2008). Overall, the previous research findings on the provision of guidance in museum settings are inconclusive, which implies the need for more research on this topic.

2.3. Affordances of AR in museum experiences

Museums are embracing AR technologies to enhance interactive elements in exhibits and visitor experiences. The review of research on museum-based mobile learning indicates that AR was used with sensing and location technologies to provide visitors with personalized learning experiences (Lin et al., 2021). In general, two types of MR applications are used in the field of education (Lindgren & Johnson-Glenberg, 2013). The first type of application is a participatory simulation where learners are situated in the system, acting as one of the components. The classic example is a virus simulation where a learner wearing a Thinking tag acts as an agent and interacts with other agents to learn about the complex algorithm of controlling disease in a dynamic simulated environment (Colella, 2000). The second type of application is an interface responsive to users' physicality and location as input; this has been expanded with advances in computing methods that can detect and process location and biological data.

Some studies have investigated the cognitive function of gestures and AR systems in the field of science education. Smith et al. (2014) used a Kinect sensor to develop a simulation-based program that allows learners to see their appearance and arm motion and confirmed the effectiveness of learning the concept of angle through this gesture-based simulation. Their study demonstrates that effective conceptual learning can be achieved when learners understand meaning by linking physical movement with the visual image on the screen and can explain concepts in connection with personal experiences. Han and Black (2011) developed a simulation for mental-model learning based on the idea that learning with simulation including movement and animation can be effective for complex learning. Simulation programs that integrate learner movement and animation provide perceptually-enhanced learning experiences. Johnson-Glenberg et al. (2014) developed an AR learning environment in that learners can perform various chemical experiments through their body movements and found that learners could learn more effectively through sensory experiences and physical movement than through static learning.

However, the dominant use of XR technologies in museums thus far has been the creation of virtual museums, virtual tours, and augmented guides. A few studies have demonstrated empirical evidence of AR technologies on cognitive learning, mainly due to the limitations in evaluation methods and instruments. Among the few research studies available, the study by Yoon et al. (2013) is relevant to understanding the complexity of AR as scaffolding for learning experiences in science museums. The researchers compared the effect of six differently scaffolded conditions (i.e., device only, digital augmentation, posted questions, collaborative groups, posted knowledge building, and recorded knowledge building) on visitors' conceptual learning. Results indicate that conceptual learning gains are high in digital augmentation, posted questions, and collaborative groups. Another interesting finding is that as scaffolds increase, the level of informal participation behaviors decreases, except in collaborative groups. They also found that digital augmentation through AR is an effective scaffold for conceptual learning.

3. Developing a gesture-based AR exhibit

3.1. Design

3.1.1. Content design

In this study, the topic “conductor resistance” was specifically chosen to develop the gesture-based AR simulation since invisible concepts such as resistance, electrons, and voltage are challenging to understand in concrete and practical terms. In traditional classrooms, teachers demonstrate these “current and resistance” concepts using an experiment with real bulbs. However, since the difference in resistance values—which depend on connection methods—is small for the voltage of the battery used in a typical experiment, it is difficult to obtain experimental results that can accurately compare subtle differences.

From the perspective of embodied cognition, human gestures have a metaphorical meaning, and the exhibit design should consider the relevance of metaphors conveyed by specific gestures. However, the more complicated the metaphorical meaning, the more likely it is that there will be differences between the designer’s intention and the visitors’ interpretation. Action-concept congruency, hence, is essential to help visitors easily discover metaphorical meanings between bodily movement and cognitive mapping. Accordingly, we intended to design a gesture-based AR exhibit that delivers a high level of action-concept congruency.

Table 1 presents the relationships between actions by a user, representations in the exhibit, and concepts as learning content that we designed to achieve this action-concept congruency. Conductor resistance is a property of a conductor defined as the amount of opposition to the flow of electric current through a conducting medium. The resistance of a conductor is proportional to its length and inversely proportional to its cross-sectional area. To imply this relationship through gestures, the length of the conductor is changed by users’ horizontal hand movements, and the cross-sectional area of the conductor is changed by users’ vertical hand movements.

Table 1. The relationship between actions, representations, and concepts

Action (User)	Representation (Exhibit)	Concept (Learning content)
Hand movements in the single-player mode • Hold still • Move hands horizontally (related to X-axis) • Move hands vertically (related to Y-axis)	Change of • Length and thickness of the conductor • Brightness of the bulb • Numerical data of the resistance and the current intensity	• Single resistance control by variation of the conductor • Correlation with the brightness of the bulb and the current intensity • Correlation with the current intensity and the resistance
Hand movements in the first two-player mode • Hold still • Move hands horizontally (related to-X axis) • Move hands vertically (related to Y-axis)	Change of • Length and thickness of each conductor • Brightness of the bulb in the connected circuit • Numerical data of the resistance and the current intensity of the connected circuit	• Composite resistance in serial connection control by variation of the multi-conductors
Hand movements in the second two-player mode • Hold still • Move hands horizontally (related to X-axis) • Move hands vertically (related to Y-axis)		• Composite resistance in parallel connection control by variation of the multi-conductors

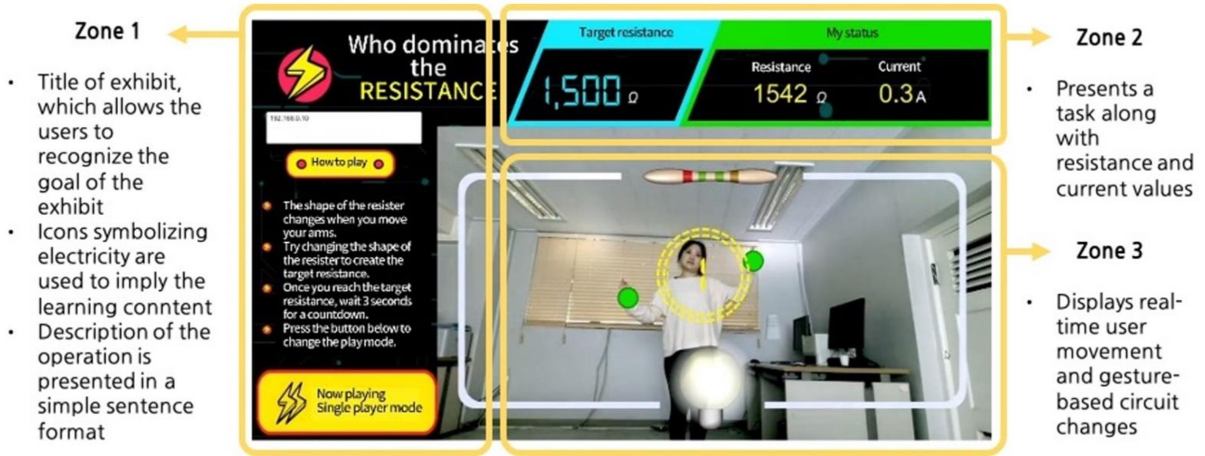
3.1.2. Design principles

A set of design principles were applied when developing the interface and tasks in the gesture-based AR exhibit to enhance user experiences and to promote conceptual learning through gestures and immersion. Here, we discuss three key design principles drawn from the literature on interaction design, embodied cognition, AR, and interactive exhibits (e.g., Antle et al., 2008; Dunleavy, 2014; Gutwill & Allen, 2012; Perry et al., 2008; Preece et al., 2015) that guided the design of our gesture-based AR exhibit system: (a) mirror-type display, (b) visual feedback, and (c) gamified collaborative tasks.

First, with a mirror-type display and simplified structure, we designed the affordance of the exhibit to be perceptually obvious to users so they understand how to interact and what to interact with. A mirror-type display allows users to recognize their images in real-time and perceive their gestures using balance and eyesight, which can promote instant participation (Park et al., 2016). Also, the control of the exhibit is visible with simple visual objects as indicators of the user's hand movements, which enables the user to intuitively grasp the meaning through those hand movements.

Second, we used a visual feedback mechanism and clear division of display areas to help users observe what actions were taken and completed (Perry et al., 2008; Preece et al., 2015). As presented in Figure 1, the display of individual users is divided into three zones. The title of the exhibit and instructions are shown in Zone 1. The icons symbolizing electricity are used to imply the learning content, and the instructions are presented in simple sentences. Zone 2 presents tasks, and various states such as individual resistance, total resistance, individual current intensity, and total current intensity that is displayed as numerical data. In Zone 3 that occupies the largest portion, users can see changes in their movements.

Figure 1. Interface design



Further, users can visually confirm the shape and size of the resistor and the change in bulb brightness resulting from their gestures. A two-dimensional virtual circuit graphic representing the change of resistance and current is overlaid on the user image drawn from the sensor. This allows users to recognize gestures (changes in hand movements) more easily. The visual feedback when matching a target resistance is overlaid on the screen. A countdown starts at ±15% of the target resistance value to help users intuitively identify how to move hands to match the target resistance value.

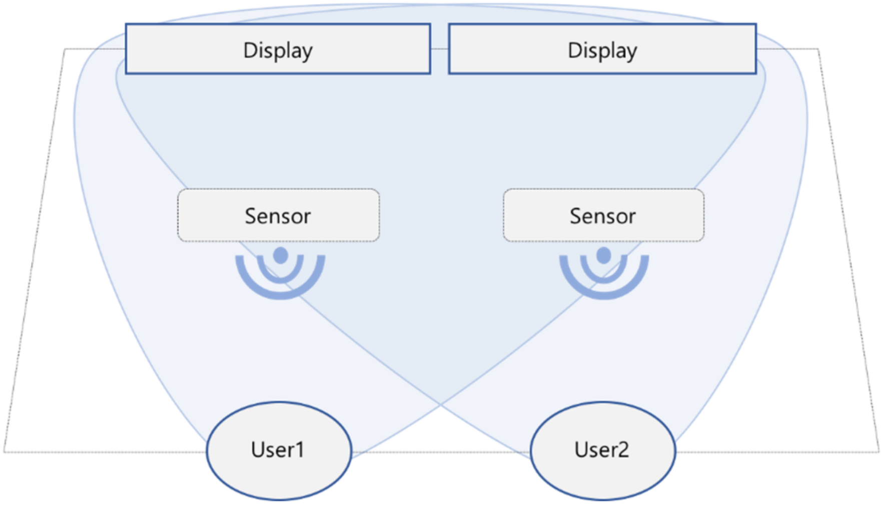
Third, the complexity in the experience structure was gradually increased with the gamified collaborative tasks to reduce cognitive load and to promote social interaction (Dunleavy, 2014). We maintained consistency in the interface design between single-player and two-player modes. Since the single-player mode contains relatively low-level concepts, the user can expend reduced cognitive load on learning basic principles involved in operating the exhibit and focus on relatively high-level concepts in the two-player mode. The tasks were also designed with gamification such as challenges, scores, and rewards. The collaborative task in the two-player mode was expected to promote more social interaction, which is an effective strategy to increase complex conceptual understanding (Horn et al., 2012: Yoon et al., 2013).

3.2. Development

The Kinect sensor was used to recognize and process users' motions. The sensor accuracy for detecting hands gesture was examined by several research studies that reported lower error rates of the Kinect sensor for gait and posture analysis (Clark et al., 2019; Ren et al., 2013). The simulation was developed using the Unity engine. An individual display was presented on the screen located in front of each user. Figure 2 illustrates the gesture-based AR exhibit design from a bird's-eye view.

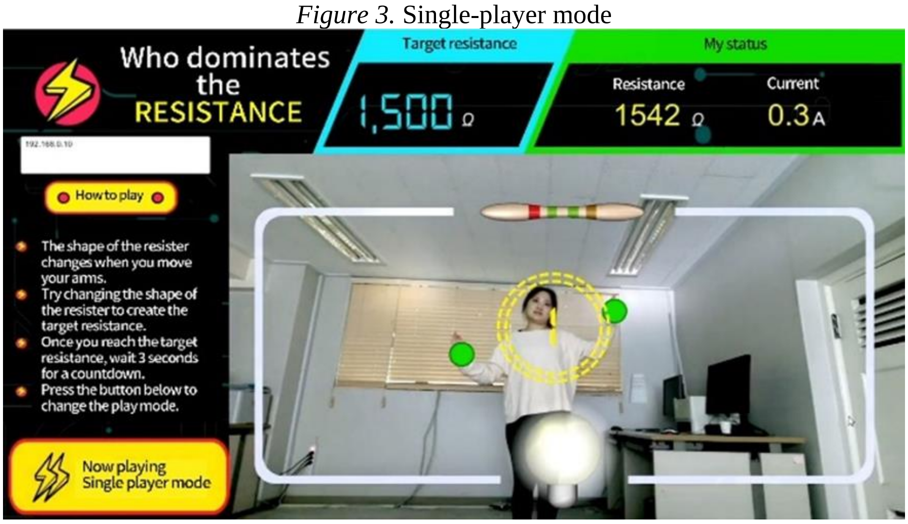
The designed scenario engages users in an interactive simulation to see the connection between their physical movement and the digital display. When visitors enter the exhibit area and stand in front of the sensor, the sensor starts detecting the visitors' gestures—hands movements in particular. The image of the visitors and the electric circuit appears on the screen. The visitors' hands movements change the length and thickness of the conductor that determines the resistance. Visitors can see the subsequent change of resistance values and current intensity through numerical data and bulb's brightness on the screen.

Figure 2. Gesture-based AR exhibit design: a macro view



The exhibit supports both single-player and two-player modes. In the single-player mode (see Figure 3), a user receives tasks to match specific resistance values. The task does not have a clear start or endpoint and is maintained until a target resistance value, given randomly, is achieved by users. When the user maintains a target value for three seconds, feedback on a successful attempt is provided, and a new value is given to the user. In the two-player mode, dyads are given the task of matching a target value collaboratively. Each user's gestures control the individual resistance values in the "serial connection" and "parallel connection" so that each user constitutes a circuit as a single resistor. In this case, the total resistance of the entire circuit needs to be achieved, not just the resistance of the individual circuit.

Figure 3. Single-player mode



4. Experiment method

4.1. Participants

After developing the gesture-based AR exhibit, we moved to the next phase to investigate its effect on complex conceptual learning in differently guided conditions. The research question examined was "is there any significant difference between guided and unguided experiences in terms of the effect on complex conceptual learning?" A quasi-experimental study was conducted with 40 middle and high school students (aged 15-17) who were recruited through convenience sampling. Since the exhibit used in this study conveys information about "current and resistance," which is taught in the ninth-grade (age 15) science curriculum in Korea, we recruited students from middle and high schools. The experiment was conducted in a lab setting due to the technical and logistical difficulty of installing the developed exhibit in a science museum. To comply with the ethics of human subject research, we explained the purpose and methods of the study before the experiment and obtained the informed consent form from all participants. They were voluntary and each received 20,000 won (about USD 20) gift card as an incentive for their participation.

4.2. Experiment design and procedures

In the experiment, participants were randomly assigned to one of two conditions: guided (n = 20) and unguided (n = 20). Table 2 shows the distribution of participants for each condition. The whole experiment lasted for two weeks and each session was about one-hour long.

Table 2. Experimental conditions

		Guided condition (n = 20)	Unguided condition (n = 20)	Total (n = 40)
Gender	Male	8	8	16
	Female	12	12	24

The guided group received explanatory visual guidance before interacting with the exhibit and verbal guidance by a docent during interacting with the exhibit. We selected verbal guidance by a docent rather than integrated on-screen guidance for two reasons. First, verbal guidance by a docent is one of the most commonly used types of guidance provided in museum learning settings. Second, our design intended to minimize the amount of textual information in the system so that users could focus on the augmented information mediated by their gestures. Table 3 presents the content of guidance presented to the guided group. Before the participants started interacting with the exhibit, the docent provided visual guidance of an electric circuit and two types of connection. During the single-player mode, the docent explained low-level concepts. When the participants switched to the two-player mode, the docent provided verbal guidance related to higher-level concepts. The unguided group did not receive any explicit guidance related to the concepts embedded in the exhibit. Only brief instructions about manipulating the exhibit were provided to the unguided group.

Table 3. Sample content provided to the guided group

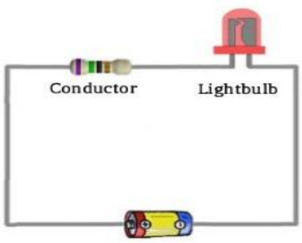
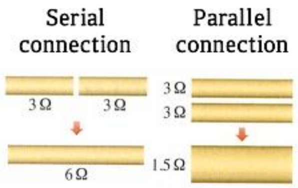
Level	Visual guidance	Verbal guidance
Low-level concepts		• Current is the rate of flow of electric charges that travel along an electrical circuit. • Resistance is a measure of the opposition to current flow in an electrical circuit. • If the material type of resistor is constant, the resistance is proportional to the length of the resistor and inversely proportional to the cross-sectional area.
High-level concepts		• If the voltage is constant, current and resistance are inversely proportional. • When several resistors are connected in series, the composite resistance becomes larger. • When several resistors are connected in parallel, the composite resistance becomes smaller.

Figure 4 presents the overall experiment procedure. A pre-test was conducted to measure the level of participants' prior knowledge about the topic. During the experiment, the participants interacted with the exhibit for 15 minutes in both single-player and two-player modes. The exhibit was arranged to be used by a dyad standing side-by-side and viewing the screen projected on the wall (see Figure 5). The exhibit first presented the single-player mode in serial connections, and then the two-player mode in parallel connections. In the single-player mode, the main tasks given to each user were to achieve a target value by changing the length and thickness of the conductor with their gestures. In the two-player mode, since two conductors were connected into one circuit, the dyad needed to collaborate in a problem-solving process. After using the exhibit, the participants took a post-test that included the same items as in the pre-test. The items were re-ordered and wordings were slightly modified to reduce the learning effect of repeated measures.

Figure 4. Experiment procedure

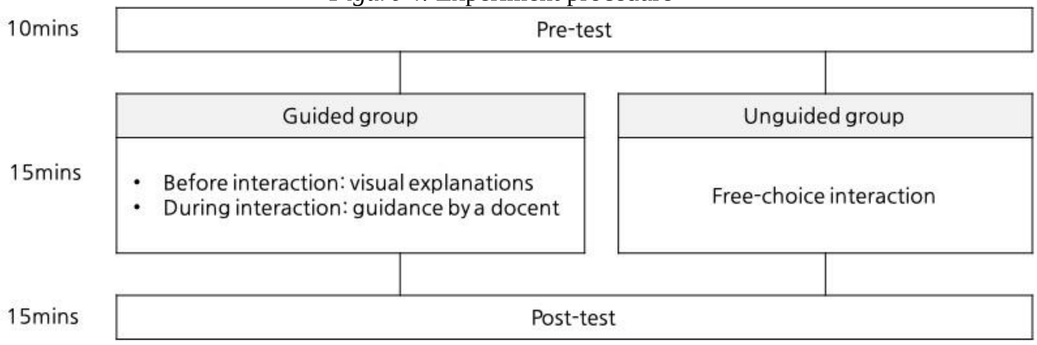
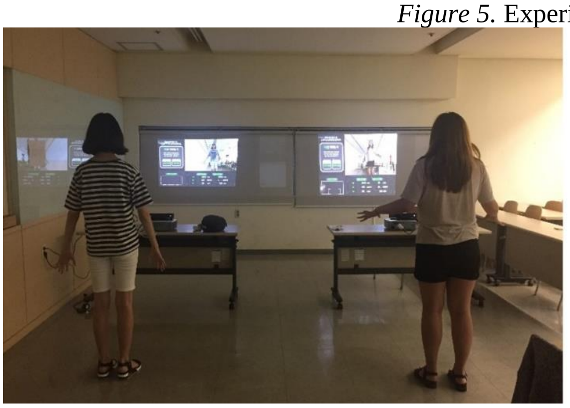


Figure 5. Experiment setup



Projectors
Image Brightness: 3,300lumens, Image Contrast Ratio: 30,000, Resolution: 800*600(SVGA)

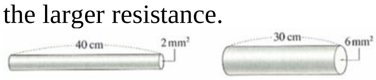
Computers
Processor: Intel Core i5, Graphics adapter: Intel HD Graphics 520, Connections: USB 3.0(with the sensor), Display Port(with the projectors), Operating System: Microsoft Windows 10 64bit

Sensors
Kinect for Windows, Developer: Microsoft, Type: Motion controller

4.3. Data collection and analysis

We assessed the cognitive learning effect with multiple-choice questions (see Table 4). The test included 16 questions, with eight items measuring low-level concepts (e.g., abstract concepts) and eight items measuring high-level concepts (e.g., relational principles). Here, we applied Gagne's (1965) hierarchical learning theory to classify the level of test items. According to Gagne, learning tasks for intellectual skills can be organized in a hierarchy according to cognitive complexity. Concept learning involves learning abstract concepts that do not have concrete physical characteristics. Principle learning, on the other hand, is connecting two or more concepts. While low-level questions measure the understanding of individual concepts and simple principles, higher-level questions measure relationships between multiple concepts and complex principles.

Table 4. Sample test items

Cognitive level	Type of learning	Concept/Principles	Sample item
Low-level	Single concepts, Simple principle	• Current • Resistance • Conductor • Conservation of charge • Resistance of a conductor	Considering that the two conductors are of the same material, choose which will make the larger resistance. 
High-level	Relational concepts, Complex principles	• Relationship between current and resistance • Series connection of resistors • Parallel connection of resistors	Which of the following is true regarding the relationship between resistance and current? (a) The larger the resistance, the larger the current. (b) The larger the resistance, the smaller the current. (c) Even if the resistance increases, the strength of the current does not change.

The questions were based on the science textbook published in the Korean Ministry of Education. To verify the content validity, one middle school female science teacher reviewed the test items with two students in her class. The teacher had a degree in Physics education and four years of teaching experience in a middle school. Since she collaborated with the research team on the design of the AR exhibit, she was able to evaluate the congruency between the test items and the exhibit design. Necessary changes, which were mostly changes in wording, were made based on their comments.

Test data were analyzed using SPSS. The Shapiro-Wilk test was conducted for checking normality. The normality at the pre-test was .56 while the post-test was .04. The Shapiro-Wilk test results showed that while the normality assumption was met for the pre-test data, the post-test data significantly deviated from the normal distribution. Hence, we used nonparametric statistics that do not require the normality assumption. The Wilcoxon signed-rank test was used to determine if the treatment affected test scores. The Mann-Whitney U test was used to examine the difference between the guided and unguided groups. The Mann-Whitney U test is suitable for testing the differences between two independent groups and is used with continuous scale data when normality cannot be assumed (Noh, 2015). The probability of significance (p-value) was set at .05.

5. Results

5.1. Comparison before and after exhibit experiences

To test the overall cognitive learning effect of the gesture-based AR exhibit, we examined the change in test scores before and after the exhibit experiences. Table 5 presents the descriptive statistics according to the level of item difficulty. The overall pattern indicates that the participants improved from the pre-test to the post-test for both low-level and high-level questions.

Table 5. Descriptive statistics (n = 40)

Level	Test	Mean	*SD*	Min-Max (0-16)
All	Pre-test	7.95	3.41	0–15
	Post-test	11.23	3.05	5–16
High-level	Pre-test	3.15	2.17	0–8
	Post-test	5.45	2.01	0–8
Low-level	Pre-test	4.50	1.73	0–8
	Post-test	5.78	1.62	2–8

As shown in Table 6, the Wilcoxon signed-rank test analysis revealed that the differences from the pre-test to the post-test were statistically significant. Based on the negative rank, all questions were z = –5.29 (p < .05), high-level questions were z = –4.88 (p < .05), and low-level questions were z = –4.03 (p < .05). Overall, the results show that the participants had significantly improved cognitive conceptual understanding after experiencing the gesture-based AR exhibit.

Table 6. Wilcoxon signed-rank test results pre-test to post-test (n = 40)

		N	Average rank	Sum of ranks	*z*	*p*
All (post-pre)	Negative ranks	1[a]	12.50	12.50	–5.29	.00*
	Positive ranks	38[b]	20.20	767.50		
	Ties	1[c]				
	Total	40				
High-level (post-pre)	Negative ranks	10[d]	15.70	157.00	–4.88	.00*
	Positive ranks	25[e]	18.92	473.00		
	Ties	5[f]				
	Total	40				
Low-level (post-pre)	Negative ranks	7[g]	9.29	65.00	–4.03	.00*
	Positive ranks	27[h]	19.63	530.00		
	Ties	6[i]				
	Total	40				

Note. [a] Post-test < Pre-test, [b] Post-test > Pre-test, [c] Post-test = Pre-test, [d] Post-test < Pre-test, [e] Post-test > Pre-test, [f] Post-test = Pre-test, [g] Post-test < Pre-test, [h] Post-test > Pre-test, [i] Post-test = Pre-test, $^*p < .05$.

5.2. Comparison by guidance condition

Table 7 presents the Mann-Whitney U test results of cognitive learning effects according to the guidance condition. Overall, the unguided condition showed higher post-test scores than the guided condition for both low-level and high-level questions. However, the differences between the two groups were not statistically significant: $z = –1.31$ ($p > .05$) for all questions, $z = –.94$ ($p > .05$) for high-level questions, and $z = –1.29$ for low-level questions ($p > .05$).

Table 7. Mann-Whitney U test results of post-test by the guidance condition ($n = 40$)

	Guided ($n = 20$)			Unguided ($n = 20$)			Mann-Whitney U	z	p
	$M \pm SD$	Average rank	Total rank	$M \pm SD$	Average rank	Total rank			
All	10.55 ±3.36	18.10	362.00	11.90 ±2.63	22.90	458.00	152.00	–1.31	.19
High-level	5.15 ±2.23	18.80	376.00	5.75 ±1.77	22.20	444.00	166.00	–.94	.35
Low-level	5.40 ±1.64	18.15	363.00	6.15 ±1.57	22.85	457.00	153.00	–1.29	.21

Note. $^{*}p < .05$.

6. Discussion

6.1. Implications of the main findings

The purpose of this study was to develop and evaluate a gesture-based AR exhibit for complex conceptual learning. In this section, we discuss the main findings and their implications. First, the gesture-based AR exhibit used in this study showed a significantly positive effect on the cognitive learning of scientific concepts. Regarding the item difficulty level, the test scores of both high-level questions and low-level questions significantly increased after interacting with the exhibit. This result is consistent with the study by Yasar and Gurel (2016), which found the relationship between the design elements and the learning content of exhibits influences learning effects. We attribute the positive gain to the careful consideration of the relationship between scientific concepts and intended actions in the gesture-based AR exhibit used in this study. For instance, the use of horizontal and vertical gestures was intentionally designed to correspond to a visual analogy for the change in the resistor's size.

Second, concerning the effect of differently-guided conditions, this study reveals that the provision of extra guidance does not necessarily lead to significantly improved conceptual learning. Our finding differs from the previous studies that reported the positive effect of structured guidance for cognitive learning (e.g., Grenier, 2009; Hohestein & Tran, 2007; Hooper-Greenhill, 1999; Yoon et al., 2013). One possible explanation is related to the intensity of verbal interaction for each condition. Our observation revealed that dyads in the guided group mostly used gestures without verbal interactions, whereas dyads in the unguided group appeared to have more frequent verbal interactions. It is possible that the guidance provided by the docent did not significantly enhance conceptual understanding during the exhibit interaction, but rather the meaning-making through verbal interaction in dyads was a more significant factor that influenced conceptual learning processes.

Overall, this research provides important findings of the efficacy of structuring museum experiences. With the increasing concern about cognitive load in immersive systems with AR/VR technologies, several design strategies have been proposed to reduce the potential cognitive load, such as providing explicit scaffolds, avoiding textual information, and providing video narrations (Matuk, 2016). This study suggests that in interactive exhibits that allow free-hand gestures and bodily movement, additional devices and human guidance (e.g., docents) to reduce cognitive load may not be necessary or effective if an exhibit is designed with careful consideration of concepts and gestures that allows visitors to easily recognize and initiate an interaction with an exhibit. Furthermore, methods for how to support collaborative meaning-making processes among visitors should be an important consideration when designing XR systems in museum settings. This is consistent with the finding by Yoon et al. (2013) on the efficacy of collaborative activities for conceptual learning in a science museum.

6.2. Limitations and areas for future research

We discuss some limitations of our study and, and suggestions for future research as potential solutions to each limitation. First, since this study was conducted with a small sample size under a controlled lab setting, the generalization of findings may be limited to similar contexts. Future research should examine more natural visitor experiences with a larger sample size in museum settings such as how visitors perceive their visit and remember their experiences after leaving the museum. Second, this study did not consider various personal and contextual factors such as the purpose of visiting a science museum, interest in science, and preferred exhibit types that might have influenced interactions and gesture patterns. Third, this study did not include a control group where users experience a non-interactive exhibit, due to the difficulty of recruiting sufficient participants and controlling multiple variables. Future research should be designed with a control group to further validate the empirical evidence of interactive AR exhibits. Forth, this study used the Kinect sensor to detect users' movements. While Kinect has been reported to have a fair accuracy for detecting large movements (e.g., moving arms), we do not rule out the possibility of detection failures in small and/or fast movements. Future research may consider using more accurate sensors for detecting gestures. Lastly, this study did not analyze verbal interaction in dyads since the main focus was on the analysis of conceptual understanding. Considering that the participants often showed collaborative discourse to solve tasks, subsequent research through discourse analysis would be useful to examine the effects of verbal interactions together with gesture usage patterns.

7. Conclusion

This study makes a valuable contribution to the design of gesture-based AR exhibits by providing empirical evidence on complex conceptual understanding, which is an under-researched area. Also, this paper presents a detailed explanation of the design principles of gesture-based AR exhibits that informs the development of similar XR systems. We suggest three takeaways from the key research findings in this study: (a) the provision of extra guidance in our gesture-based AR exhibit does not necessarily lead to improved conceptual understanding, (b) the efficacy of gesture-based AR exhibits is enhanced when the design embeds guidance that reduces potential cognitive load and increases action-concept congruency, and (c) it is important to engage and promote social interaction through game elements and challenging tasks to enhance conceptual understanding and collaborative meaning-making. We believe that these findings can be used to advance the larger research discourse on XR and the design of immersive learning environments by providing empirical evidence of gesture-based exhibits with AR.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Government (NRF-2015R1C1A1A02037692).

References

Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education, 88*, 17-33.

Allen, S., & Gutwill, J. (2004). Designing with multiple interactives: Five common pitfalls. *Curator: The Museum Journal*, *47*(2), 199-212.

Antle, A. N., Corness, G., & Droumeva, M. (2008). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, *21*(1-2), 66-75.

Baydas, O., Karakus, T., Topu, F. B., Yilmaz, R., Ozturk, M. E., & Goktas, Y. (2015). Retention and flow under guided and unguided learning experience in 3D virtual worlds. *Computers in Human Behavior*, *44*, 96-102.

Braund, M., & Lelliott, A. (2017). Opening up the dialogic Space. Using questions to facilitate deeper informal learning. In *Preparing Informal Science Educators* (pp. 561-574). Springer, Cham.

Clark, R. A., Mentiplay, B. F., Hough, E., & Pua, Y. H. (2019). Three-dimensional cameras and skeleton pose tracking for physical function assessment: A Review of uses, validity, current developments and Kinect alternatives. *Gait & Posture, 68*, 193-200.

Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences, 9*(4), 471-500.

Dunleavy, M. (2014). Design principles for augmented reality learning. *TechTrends, 58*(1), 28-34.

Falk, J. H. (1997). Testing a museum exhibition design assumption: Effect of explicit labeling of exhibit clusters on visitor concept development. *Science Education, 81*(6), 679-687.

Fang, Y., Wang, K., Cheng, J., & Lu, H. (2007). A Real-time hand gesture recognition method. In *Multimedia and Expo, 2007 IEEE International Conference on* (pp. 995-998). IEEE.

Freeman, A., Adams Becker, S., Cummins, M., McKelroy, E., Giesinger, C., & Yuhnke, B. (2016). *NMC Horizon Report: 2016 Museum edition*. The New Media Consortium.

Gagne, R. M. (1965). The Learning of concepts. *The School Review, 73*(3), 187-196.

Gardner, R. (1995). *Make an interactive science museum: Hands-on exhibits*. Tab Books.

Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2018). Applications of Augmented Reality in informal science learning sites: A Review. *Journal of Science Education and Technology, 27*(5), 433-447.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516-522.

Grenier, R. S. (2009). The Role of learning in the development of expertise in museum docents. *Adult Education Quarterly, 59*(2), 142-157.

Gutwill, J. P., & Allen, S. (2012). Deepening students' scientific inquiry skills during a science museum field trip. *Journal of the Learning Sciences*, *21*(1), 130-181.

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education, 57*(4), 2281-2290.

Hinrichs, U., & Carpendale, S. (2011). Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3023-3032). ACM.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.

Hohenstein, J., & Tran, L. U. (2007). Use of questions in exhibit labels to generate explanatory conversation among science museum visitors. *International Journal of Science Education*, *29*(12), 1556-1580.

Hooper-Greenhill, E. (1999). *The Educational role of the museum*. Routledge.

Horn, M., Atrash Leong, Z., Block, F., Diamond, J., Evans, E. M., Phillips, B., & Shen, C. (2012, May). Of BATs and APEs: An Interactive tabletop game for natural history museums. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2059-2068). ACM.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, *106*(1), 86-104. https://doi.org/10.1037/a0034008

King, H., & Tran, L. (2017). Facilitating deep conceptual learning: The Role of reflection and learning communities. In *Preparing Informal Science Educators* (pp. 67-85). Springer, Cham.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An Analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Korn, R., & Jones, J. (2000). Visitor behavior and experiences in the four permanent galleries at the tech museum of innovation. *Curator, 43*(3), 261-281.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist, 42*(2), 109-113.

Lin, C. J., Tang, K. Y., & Tu, Y. F. (2021). Advancements and research topics of museum-based mobile learning: A Review of journal publications from 2008 to 2019. *Interactive Learning Environments*, 1-24. https://doi.org/10.1080/10494820.2021.1981394

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, *42*(8), 445-452.

Lyons, L., Slattery, B., Jimenez, P., Lopez, B., & Moher, T. (2012). Don't forget about the sweat: Effortful embodied interaction in support of learning. In *Proceedings of the 6th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 77-84). ACM.

Matuk, C. (2016). The Learning affordances of augmented reality for museum exhibits on human health. *Museums & Social Issues*, *11*(1), 73-87.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99-113.

Noh, H. S. (2015). *Non-parametric statistical analysis and correspondence analysis using SPSS*. Jiphill Media.

Park, H. S., Kim, H. W., & Park, C. J. (2016). Dynamic-interaction UI/UX design for the AREIS. In *International Conference on Human-Computer Interaction* (pp. 412-418). Springer, Cham.

Perry, J., Klopfer, E., Norton, M., Sutch, D., Sandford, R., & Facer, K. (2008, June). AR gone wild: Two approaches to using augmented reality learning games in Zoos. In *Proceedings of the 8th international conference on International conference for the learning sciences* (Vol. 3, pp. 322-329). International Society of the Learning Sciences.

Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design* (4th ed.). Wiley.

Ren, Z., Yuan, J., Meng, J., & Zhang, Z. (2013). Robust part-based hand gesture recognition using Kinect sensor. *IEEE Transactions on Multimedia, 15*(5), 1110-1120.

Shaby, N., Ben-Zvi Assaraf, O., & Tal, T. (2019). An Examination of the interactions between museum educators and students on a school visit to science museum. *Journal of Research in Science Teaching, 56*(2), 211-239.

Smith, C. P., King, B., & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior, 36*, 95-108.

Stronck, D. R. (1983). The Comparative effects of different museum tours on children's attitudes and learning. *Journal of Research in Science Teaching, 20*(4), 283-290.

Sun, J. C. Y., & Yu, S. J. (2019). Personalized wearable guides or audio guides: An Evaluation of personalized museum guides for improving learning achievement and cognitive load. *International Journal of Human–Computer Interaction, 35*(4-5), 404-414.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tobias, S., & Duffy, T. M. (2009). The Success or failure of constructivist instruction: An Introduction. In *Constructivist Instruction* (pp. 15-22). Routledge.

Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science, 322*(5902), 682-683.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636.

Wilson, R. A., & Foglia, L. (2017). *Embodied cognition, The Stanford encyclopedia of philosophy* (Spring 2017 ed.), Edward N. Zalta (Ed.). https://plato.stanford.edu/archives/win2021/entries/embodied-cognition

Yasar, E., & Gurel, C. (2016). Science museum exhibits' summative evaluation with knowledge hierarchy method. *European Journal of Physics Education, 7*(1), 25-36.

Yoon, S. A., Elinich, K., Wang, J., Van Schooneveld, J. B., & Anderson, E. (2013). Scaffolding informal learning in science museums: How much is too much? *Science Education, 97*(6), 848-877.

Factors Influencing University Students' Intention to Engage in Mobile-assisted Language Learning through the Lens of Action Control Theory

Hui-Tzu Hsu[1*] and Chih-Cheng Lin[2]
[1]Language Centre, National Chin-Yi University of Technology, Taichung, Taiwan (R.O.C.) // [2]Department of English, National Taiwan Normal University, Taiwan (R.O.C.) // lindahsu85@gmail.com // cclin@ntnu.edu.tw
[*]Corresponding author

(Submitted October 6, 2021; Revised December 18, 2021; Accepted January 14, 2022)

ABSTRACT: Mobile technology is regarded as a helpful tool facilitating language learning. However, the success of mobile technology largely depends on learners' acceptance. This study explored the factors that may affect students' intention formation regarding mobile-assisted language learning (MALL) in the context of higher education through the lens of action control theory. The study adopted mixed methods: an online survey of 557 students and individual interviews with 70 students. The findings indicated factors in each of the three dimensions (preoccupation, hesitation, and volatility) of action control theory that positively or negatively influenced the students' intention to use mobile technology for language learning. According to the findings, these influential factors may be related experiences in the preoccupation dimension, design and feature interference of MALL applications and teachers' teaching style influence in the hesitation dimension, and overall appraisal and performance impact and other novelty interference in the volatility dimension. Students' success in initiating and completing a MALL task depends on mainly depends on their acceptance of MALL, and this acceptance is affected by these factors in a positive or negative direction. The strengthening of the positive influence and the weakening of the negative influence caused by these factors should be paid attention to in the process of performing and engaging in a MALL task. Students' concerns regarding the use of mobile technology in language education are addressed with suggestions for future research and practice in light of the findings.

Keywords: Mobile technology, Language learning, Learning intention, Action control, Higher education

1. Introduction

Technology-enhanced learning, in particular with mobile devices (e.g., smartphones and tablets), is under the spotlight in higher education. The majority of mobile device users are college students aged 18–29 years (Pew, 2021; Poushter, 2016). Advancements in mobile technology have greatly supported language learning in higher education settings (Crompton & Burke, 2018; Hwang & Fu, 2019; Ke & Hsu, 2015; Kukulska-Hulme et al., 2017; Reinders & Pegrum, 2017) because such technology can be a tool conducive to fostering autonomous language learning anywhere and at any time (Reinders & Benson, 2017) and enhance learning outcomes, interaction, and positive perceptions (Golonka et al., 2014). Due to the benefits of mobile-assisted language learning (MALL), MALL technology has gradually been integrated into language curriculum; however, factors positively or negatively influencing the integration of MALL cannot be ignored (Chwo et al., 2016).

The success of implementing mobile technology for language learning depends on student acceptance (Hsieh et al., 2017). The previous studies adopted the technology acceptance model (TAM), the extended TAM (e.g., the General Extended Technology Acceptance Model for E-learning, Abdullah & Ward, 2016), self-determination theory of motivation (Nikou & Economides, 2017), and activity theory (Lin et al., 2020) to investigate factors influencing learners' intention to adopt mobile learning or MALL. Unlike the previous studies, the present study adopted action control theory (ACT; Kuhl, 1994a) to design survey items for each of the three dimensions (preoccupation, hesitation, and volatility) based on the context of MALL to investigate possible factors influencing learner intention transformation to accept or reject MALL. Limited research has focused on behavioural intention changes during MALL according to a meta-analysis conducted by Hwang and Fu (2019). The causes underlying behavioural intention change in MALL are poorly understood. In addition, in second language acquisition (SLA) research, ACT proposed by Kuhl (1994a), a prominent theory in mainstream psychology, has been adopted to investigate foreign language learner behaviour and motivation (Dörnyei & Ryan, 2015; Ellis & Shintani, 2014; Khany & Amiri, 2018; MacIntyre & Blackie, 2012; MacIntyre & Doucette, 2010). ACT focuses on continuums of two poles (state and action orientations) in three dimensions (preoccupation, hesitation, and volatility). These tendencies may affect whether an individual is able to initiate, focus, and follow up on a task (Jaramillo et al., 2007). ACT has high explanatory potential for outcome variables in the context of second language learning (Ellis & Shintani, 2014).

The current study adopted ACT to explore possible factors affecting behavioural intention changes in the course of MALL tasks in higher education settings. In view of this, the study aimed to explore what determinants could influence intention transformation when learners were engaging in MALL tasks in the three dimensions (preoccupation, hesitation, and volatility) of ACT and to deeply investigate learners' perception that the interference factors could influence their intention changes to adopt MALL in a positive or negative direction. The study contributes to the field of MALL use in higher education in the following respects. First, the investigation may broaden knowledge of the causes underlying behavioural intention change in MALL tasks. Second, the findings can provide a reference for higher education educators and developers of MALL systems to tailor tasks promoting positive change in student intention without interference from negative factors that severely hamper the completion of MALL tasks. Finally, to our knowledge, this is the first study to adopt a cross-sectional survey and individual interviews to examine the factors influencing student intention transformation in MALL tasks in higher education. In the following sections, previous studies related to this study in Background, research participants, designs, and analytical methods in Research Methodology, and research findings and discussion in Results and Discussion provide a comprehensive picture of this study.

This study investigated possible factors affecting learners' behavioural intention to accept MALL through the lens of ACT. We addressed two research questions:

- RQ1. What factors affect the behavioural intention of university students to use a mobile device for language learning through the lens of ACT?
- RQ2. What perceptions do students have regarding the factors affecting their MALL intention through the lens of ACT?

2. Background

2.1. Mobile-assisted language learning

Handheld mobile technologies are influencing how higher education students learn languages. Mobile devices now outnumber conventional desktops (Pegrum, 2014). Interest in using mobile devices such as smartphones and tablets for educational objectives (Duman et al., 2015) has gradually increased. Mobile learning is defined as learning involving the use of a mobile device (Crompton, 2013). In a special issue of *ReCALL* in 2008, Kukulska-Hulme and Shield introduced MALL as a developing field. In MALL, learners adopt mobile technology to engage in language learning tasks (Burston, 2015; Kukulska-Hulme et al., 2017; Shadiev et al., 2017).

In higher education environments, many researchers have highlighted the role of mobile technology in effective language learning. Mobile technology can enhance and improve language learning performance because of its key advantages, including ubiquity, adaptability to personal study habits, higher authenticity, easy accessibility of information, and continuity of study on different devices (Burston, 2015; Duman et al., 2015; Kukulska-Hulme et al., 2017; Loewen et al., 2019; Petersen & Sachs, 2016; Reinders & Pegrum, 2017). Students can obtain materials and study languages anytime and anywhere. Numerous studies have indicated that mobile technology could become a helpful tool for vocabulary learning (Chen & Chung, 2008; Chen et al., 2019; Kim, 2011; Lin & Lin, 2019; Ono et al., 2015; Stockwell, 2007; Wu & Huang, 2017), speaking and listening skills development (Hwang & Chen, 2013; Hwang et al., 2014; Nguyen et al., 2018), reading ability enhancement (Chen & Hsu, 2008; Hsu et al., 2013), grammar learning (Li & Hegelheimer, 2013; Wang & Smith, 2013), and writing skills improvement (Jiang & Zhang, 2020).

MALL researchers have identified drawbacks of mobile technology and factors affecting acceptance of the technology for language learning. Studies have indicated that small screen size, typing problems, connection problems, and privacy intrusion can negatively influence learning intention and outcomes (Lai & Zheng, 2018; Li & Hegelheimer, 2013; Thornton & Houser, 2005). In addition, not all students are willing to use mobile learning applications (Hsu, 2015; Kim et al., 2017; Stockwell, 2010). Successful adoption of mobile technology for language learning is dependent on student acceptance of such technology (Hsieh et al., 2017). Increasingly, studies have focused on investigation of learner intention to engage in MALL tasks with the adoption of the technology acceptance model (Chang & Hsu, 2011; Chang et al., 2013; Chen, 2018; Hoi, 2020; Nie et al., 2020). Because mobile technology as a major role can be conducive to language learning, the factors affecting students' intention to adopt such technology for language learning should be investigated.

2.2. Action control theory

ACT, proposed by Kuhl (1994a), is a prominent theory in mainstream psychology. According to the theory, individual differences play a central role in emotion regulation, cognition, and behaviour in fulfilling intentional action. Action control takes the effects of mediating processes into account because these processes influence intentions. In the theory, a continuum exists between action and state orientations. Action-oriented learners can pursue their intentions or goals and then use appropriate cognitive skills to successfully convert their intentions into actions. By contrast, state-oriented learners do not have the ability to transform their intentions into actions because of negative states (e.g., a fear of failure or hesitation to initiate an action; Diefendorff et al., 2000).

Apart from the general concept of action-state orientation, ACT has three components, namely preoccupation, hesitation, and volatility. Preoccupation is defined as the degree to which individuals manage interfering and unpleasant thoughts in a past, present, or future state. The opposing poles in the preoccupation component are preoccupation versus disengagement. At the disengagement pole, action-oriented individuals disentangle themselves from thoughts of alternative targets or unpleasant events that may hamper progress in a given task. By contrast, a state-oriented individual at the preoccupation pole cannot disentangle from unpleasant experiences such as past failures. Hesitation pertains to the degree to which individuals have trouble converting decisions into action and the hesitation between continuing an already initiated task and beginning a completely new one. The opposing poles in the hesitation component are hesitation versus initiative. Action-oriented individuals at the initiative pole easily initiate work on tasks. By contrast, a state-oriented individual at the hesitation pole may not have the behavioural capacity to launch actions. Volatility is defined as the degree to which individuals are distracted when performing an already initiated action. Action-oriented learners effectively sustain strong and sustained intention until the task is completed. However, a state-oriented learner is easily distracted from current tasks, which impairs their overall performance. They may struggle to appropriately initiate new and novel tasks.

ACT, borrowed by SLA researchers, has gradually earned a place in second language (L2) motivation research (Dörnyei, 2001; Dörnyei, 2005; Ellis & Shintani, 2014) and has been empirically corroborated (Khany & Amiri, 2018; MacIntyre & Blackie, 2012; MacIntyre & Doucette, 2010; MacIntyre et al., 2001). Khany and Amiri (2018) examined a proposed model in terms of ACT, motivational self-system of L2, and learners' motivated behaviour in the context of a foreign language. They reported that preoccupation, hesitation, and volatility had negative impacts on L2 learners' motivated behaviour and learning experience. Students' ideal L2 self was positively correlated with volatility, but a negative correlation existed between students' "ought-to" L2 self and hesitation. MacIntyre and Blackie (2012) assessed the contribution of ACT in predicting nonlinguistic outcomes by using the Gardnerian socioeducational model and Pintrich's expectancy-value model. They discovered that action control is highly related to nonlinguistic outcomes. Hesitation can predict willingness to communicate, language anxiety, perceived communication competence, and the intention to continue L2 learning. Preoccupation was a determinant of perceived communicational ability, willingness to communicate, and language anxiety, and volatility was a predictor of language anxiety.

Despite these findings, little research examined the latent contribution of ACT in mobile technology–enhanced language learning contexts in higher education settings (Hsu & Lin, 2022). Researchers have extended the technology acceptance model with other factors to investigate learner acceptance of MALL (Chen, 2018; Nie et al., 2020; Park et al., 2012). However, little is known regarding the factors affecting learner acceptance of MALL in the process of behavioural intention transformation in the three dimensions (preoccupation, hesitation, and volatility) of ACT. Accordingly, this study examined possible factors in the behavioural intention transformation of university students when they are performing MALL tasks, with a focus on their perception in each dimension of ACT.

3. Research methodology

3.1. Participants and sample size

As for cross-sectional survey, the participants in this study were 557 undergraduate students of six universities in Taiwan learning English as a foreign language. All of them voluntarily filled in the online questionnaire. Of all participants, 167 (30%) were males and 390 (70%) females. Most participants (500; 89.8%) were aged under 20 years, and 57 (10.2%) participants were aged 21–25 years. Given the focus of the study on MALL, the participants' English language learning experience with mobile devices was collected to ensure that all used mobile devices to learn English. In total, 155 (27.8%) participants had less than 1 year of experience learning

English with mobile devices, with the remaining participants (402; 72.2%) having learned English with mobile devices for more than a year.

As for interviews, in all, 70 undergraduate students (56 males and 14 females) of one university in Taiwan were recruited as interviewees; 65 (92.8%) were aged ≤ 20 years and 5 (7.1%) were aged 21–25 years. They all signed the consent form and agreed to participate in the interviews of the study. The mobile devices the participants used varied: 69 (98.5%) owned a smartphone, 15 (21.4%) had a laptop, and 3 (4.2%) owned a tablet. All the interviewees had experience learning English with mobile devices. Overall, 63 (90%) had less than 1 year of English language learning experience with mobile devices, and the remaining 7 (10%) had used a mobile device to learn English for 1–3 years.

3.2. Research design

The study adopted ACT to explore the factors university students perceived as affecting their behavioural intention to use mobile devices to learn English, in the three dimensions of preoccupation, hesitation, and volatility. Mixed methods were employed for data collection: a quantitative one (a cross-sectional survey) and a qualitative one (individual interviews). Dörnyei (2007) indicated that researchers commonly employ interviews and questionnaires for data collection. A survey following online convenience sampling can be used to collect data from numerous individuals at a single point in time (Fraenkel et al., 2012). Interviews enable researchers to investigate people's views in greater depth (Kvale, 1996; Kvale, 2003).

3.3. Data collection and analysis

The survey was designed in terms of ACT to explore possible factors affecting the behavioural intention transformation process when students engage in MALL tasks, among the three dimensions (preoccupation, hesitation, and volatility). Ten open-ended questions were designed for the semi-structured interviews. To ensure content validity, the survey items and interview questions were amended a few times based on the four experts' suggestions. The major revision suggestions from the four experts were summarised as follows. First, the tense of each survey item should be revised from the past tense to future tense. Second, the four survey items in the hesitation construct should be concise and clear, not wordy. Third, the previous English learning experience should be an interference factor, so it should be one of the survey items in the preoccupation construct. Fourth, teaching style may be a possible factor interfering learner intention to initiate a MALL task, so it should be added in the hesitation construct. Fifth, in the second survey item of the volatility construct, "successful" should be used to describe MALL performance to ensure that learners will have intention to continue using mobile devices for their English learning when they have good learning performance in a MALL task. Sixth, the interview question design should be based on the ten survey items to gain an in-depth understanding of related specific events influencing the learners' intention to use mobile devices for their English learning in the three ACT dimensions. Finally, for each interview question, it should be necessary to invite interviewees to share previous related experiences and to explain whether the shared experience may positively or negatively influence their intention to use MALL in each of the three ACT dimensions. The four experts had more than 15 years of English language teaching experience. The items of the scale and interview questions were translated into Mandarin. We invited two qualified participants to fill out the Mandarin version of the questionnaire and to review the interview questions to ensure that all the items in both instruments were understandable. According to their suggestions, the items were revised to achieve the desired level of readability. The two participants who supported the pretesting process were excluded from formal sampling.

The survey comprised two sections: (1) a scale for possible factors affecting behavioural intention to continue using mobile devices to learn English, and (2) demographic information. According to Table 1, the ten scale items were designed in accordance with ACT (Kuhl, 1994a), the Action Control Scale (ACS-90, Kuhl, 1994b), and the technology acceptance model (Davis, 1989). The first three items focused on preoccupation; the following four items were related to hesitation; and the final three items measured volatility. A 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) was adopted. The ten opened-ended questions of the interviews were based on the ten items of the scale.

In the survey, 557 valid questionnaires were collected. The reliability and validity of the measured items were investigated, and Pearson correlation analysis was used to assess the relationships among the three constructs. Descriptive statistics are used to summarise and present quantitative data. Qualitative data from the interviews were analysed in three stages: description, analysis, and interpretation (Wolcott, 1994). As noted by Wolcott (1994, p. 12), description can "address what is going on here and single out some things as worthy of note and relegates others to the background;" analysis addresses "the identification of essential features and the systematic description of interrelationships among them;" and interpretation addresses "processual questions of meanings and contexts." In data analysis, recurring themes can provide in-depth information.

4. Results

4.1. Descriptive statistical analysis of the items in each construct

As presented in Table 1, among all measured items, the total average for the hesitation construct (mean [*M*] = 5.99, standard deviation [*SD*] = 1.10) was highest, followed by preoccupation (*M* = 5.64, *SD* = 1.18), and volatility (*M* = 5.51, *SD* = 1.16).

Table 1. Descriptive statistics of measured items of three ACT constructs

Construct	Measured items	*M*	*SD*	Rank
Preoccupation	1. Previous experiences of English learning through mobile devices will influence if I am willing to use mobile devices to learn English.	5.77	1.06	1
	2. Previous experiences of using mobile devices will influence if I am willing to use mobile devices to learn English.	5.59	1.24	2
	3. Previous experiences of English learning will influence if I am willing to use mobile devices to learn English.	5.55	1.24	3
Average		**5.64**	**1.18**	**-**
Hesitation	1. The interface of MALL will influence my decision on if I will use mobile devices to learn English.	5.78	1.20	3
	2. The content of MALL will influence my decision on if I will use mobile devices to learn English.	6.19	0.98	2
	3. Teachers' teaching styles will influence my decision on if I will use mobile devices to learn English.	5.74	1.29	4
	4. Convenience and usefulness of mobile devices will influence my decision on if I will use mobile devices to learn English.	6.25	0.94	1
Average		**5.99**	**1.10**	**-**
Volatility	1. My appraisal of the whole MALL procedure will influence if I am willing to continue using mobile devices to learn English.	5.95	1.04	2
	2. Successful MALL performance will influence if I am willing to continue using mobile devices to learn English.	6.15	0.94	1
	3. Other, novel things will not influence if I am willing to continue using mobile devices to learn English.	4.42	1.51	3
Average		**5.51**	**1.16**	**-**

4.2. Factor analysis

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. Varimax rotation was used to examine the construct characteristics or degree of concept. The Kaiser–Meyer–Olkin (KMO) test for sampling adequacy was used for the evaluation of size. KMO values range between 0 and 1. A high KMO value denotes several common factors between items and thus indicates suitability for factor analysis. When the KMO value is lower than 0.5, conducting a factor analysis is not recommended (Kaiser, 1974). Among the constructs, the KMO value was highest for hesitation (0.750), followed by preoccupation (0.696) and volatility (0.526; Table 2). Thus, all values were higher than the threshold value (0.5).

Table 2. KMO value of each construct

Construct	KMO value
Preoccupation	0.696
Hesitation	0.750
Volatility	0.526

4.3. Reliability and validity analysis

4.3.1. Reliability analysis

For reliability analysis (Table 3) to determine whether items had a suitable level of internal consistency, the criteria of composite reliability (CR) and Cronbach's α of at least 0.6 (Fornell & Larcker, 1981) were adopted. Latent variable values that are highly correlated indicate that the consistency of the construct reflected by the items is high. In the results, the CR of the latent variables for all constructs ranged between 0.73 and 0.80, and Cronbach's α ranged between 0.811 and 0.863, showing that the latent variables had good reliability.

4.3.2. Validity analysis

A validity analysis evaluates the correctness of a questionnaire, which is often reflected by three indicators, namely content validity, convergent validity, and discriminant validity. Again, the questionnaire items were based on relevant literature and were discussed and reviewed by four experts. Therefore, the questionnaire can be considered to have content validity. For favourable convergent validity, individual factor loadings must reach 0.5 or higher and range between 0 and 1. For favourable discriminant validity, the square root of the average variance extracted (AVE) of one construct must be larger than the correlation coefficient of the other (Fornell & Booksten, 1982; Fornell & Larcker, 1981). The convergent validity analysis revealed that individual factor loadings were all greater than 0.5, and the AVE of the latent variables ranged between 0.47 and 0.52 (Table 3). The AVE of each construct should be higher than 0.5, but we accepted 0.4 because, according to Fornell and Larcker (1981), if AVE is less than 0.5 but CR is higher than 0.6, the convergent validity of the construct is adequate. The CR of each construct was higher than 0.7; thus, an AVE between 0.4 and 0.5 was acceptable. The square roots of the AVE of the latent variables were 0.688–0.724 (Table 4). These values were acceptable (Fornell & Larcker, 1981).

4.3.3. Correlation analysis

The mean and standard deviation for hesitation were 5.99 and 1.10; those for preoccupation were 5.64 and 1.18; and those for volatility were 5.51 and 1.16. According to Table 4, preoccupation (r (555) = 0.540, $p < .01$) and hesitation (r (555) = 0.577, $p < .01$) were positively correlated with volatility. The variable preoccupation (r (555) = 0.534, $p < .01$) was positively correlated with hesitation. Generally, researchers believe that a correlation coefficient of < 0.3, 0.3–0.7, and > 0 indicates a low, moderate, and high correlation, respectively. Therefore, the three variables were moderately correlated (Hwang, 2018).

Table 3. Reliability indices and factor loadings for each construct

Construct	Measured item	Factor loading	AVE	CR	Alpha
Preoccupation	P1	0.613	0.47	0.73	0.811
	P2	0.668			0.818
	P3	0.772			0.863
Hesitation	H1	0.696	0.51	0.80	0.814
	H2	0.771			0.815
	H3	0.511			0.818
	H4	0.836			0.822
Volatility	V1	0.759	0.52	0.77	0.813
	V2	0.732			0.836
	V3	0.679			0.816

Table 4. Discriminant validity and correlation

	Preoccupation	Hesitation	Volatility
Preoccupation	0.688		
Hesitation	0.534^{**}	0.714	
Volatility	0.540^{**}	0.577^{**}	0.724

Note. $^{**}p < .01$.

4.4. Student perceptions of factors affecting MALL, in the ACT dimensions

To address the second research question, 70 interview transcripts were analysed. The steps (description, analysis, and interpretation) proposed by Wolcott (1994) were adopted to analyse the qualitative data of the interview transcripts. First, we found out the whole picture of the interview transcripts and then selected the noteworthy parts related to factors influencing learner intention changes in each of the three ACT dimensions. Secondly, we identified essential features (e.g., interference from previous experiences and application design and feature impact) and systematically analyse the interrelationship among these features. Finally, we interpreted the correlations between the features and the three dimensions (preoccupation, hesitation, and volatility) of ACT and summarised the recurring themes, offering in-depth information. The major themes that emerged from the data were as follows: related experiences in the preoccupation dimension; design and feature interference of MALL applications in the hesitation dimension; teaching style influence in the hesitation dimension; overall appraisal and performance impact in the volatility dimension; and other novelty interference in the volatility dimension.

4.4.1. Related experience in the preoccupation dimension

With preoccupation defined as interfering and unpleasant thoughts hindering individuals to initiate an act (Kuhl, 1994a), students with a positive previous experience can detach themselves from such thoughts to initiate a task, whereas those with a negative previous experience may lack such an ability. This phenomenon was noticeable in mobile-assisted English language learning. Most participants mentioned that positive experiences in mobile-assisted English language learning, using mobile devices, and English language learning may enhance their willingness to initiate a MALL task, whereas some stated that they were unable to disentangle negative experience to conduct such a task. Participant 3 reported a positive experience as follows:

I've had a good experience in adopting Quizlet to improve my English vocabulary learning. Its merits such as user-friendly interface, various learning activities, and quiz games could effectively improve my vocabulary scores. I can learn English vocabulary wherever and whenever I want. I'll initiate similar mobile-assisted learning tasks in the future because of such positive experiences.

However, Participant 15 discussed a negative experience in mobile device use for English language learning:

I don't initiate a new mobile-assisted learning task because I've had a negative English learning experience with mobile devices. My high school English teacher assigned English tasks which we needed to complete on a mobile English learning application, Duolingo. Although it was an interesting way to learn English, it made my eyes strain and some pop-up advertisements frequently interrupted my flow and learning attention. In the future, these two reasons may demotivate me to initiate such tasks.

Most participants reported that the ubiquity of mobile devices can help them easily commence English language learning tasks ("I can use my smartphone anytime and anywhere; this ubiquity can help me break the ice on mobile-assisted language learning tasks" #7). However, a few students firmly believed that their smartphone was for communication purposes rather than for English language learning ("I will not use my smartphone to engage in any English learning tasks because my smartphone is used for contacting my friends and families, rather than for English learning" #23).

Most interviewees expressed that their previous English language learning in a fixed place (e.g., classroom or cram school) caused them to resort to mobile technology because it helped them learn English anytime and anywhere (e.g., "Learning English in class or cram school did not help me effectively learn English, so I'm grateful for mobile technology which can help me learn English anytime and anywhere" #39). However, a few students reported that they preferred to learn English with teachers in person because they could obtain immediate feedback from teachers if they had questions. MALL tasks cannot offer this immediate interaction (e.g., "In past face-to-face English learning, I could immediately ask teachers questions and obtain answers in class; however, I can't ask questions and get answers immediately on MALL applications" #49). As stated earlier, students may become paralysed in the learning process because of related negative experiences or they can detach themselves from extraneous thoughts regarding unrelated goals by using positive related experiences.

4.4.2. Design and feature interference of MALL applications in the hesitation dimension

Hesitation denotes an inability to convert decisions into actions (Kuhl, 1994a). In this dimension, the perceived design and features of MALL applications may interfere with a learner's determination on whether to continue an initiated MALL task. After learners commence a MALL task, they experience the design and features of MALL application, causing hesitation between continuing the initiated task and starting a new task. Most participants reported that an intuitive and user-friendly interface design and diverse MALL features made them firmly continue initiated MALL tasks:

When I initiated a MALL task in the past, the intuitive and user-friendly interface, various and interesting learning activities, and learning anytime and anywhere could keep me engaged in the initiated task without being tempted and disturbed by other new things. (#5)

By contrast, a complex interface design, repetitive activities, and content that is too difficult or monotonous may cause learners to hesitate regarding whether to remain engaged in an initiated MALL task:

In the past, after conducting a vocabulary learning task on Quizlet for several weeks, I stopped to look for new ones because it was too boring to use repetitive vocabulary learning activities like the flash cards and match activity. (#43)

I've used the Voice of America application to enhance my listening, but I gave up using it because of its complex and unintuitive interface. (#8)

4.4.3. Teaching style influence in the hesitation dimension

Teaching style may be a primary factor influencing continual MALL. More than 90% of the participants stated that if a teacher had a vivid, interesting teaching style, they would remain engaged in MALL tasks rather than abandoning them in favour of a novel task (e.g., "One of the English teachers I had before had a vivid and interesting teaching style. She guided us to learn English words on Kahoot. I kept learning English via Kahoot because of her teaching style, not Kahoot" #54). Interviewees also indicated that traditional and demanding teaching styles could demotivate them, as indicated by the following response:

I've used mobile devices to learn English, but later I stopped using this method because of my demanding English teacher. I knew this method could be beneficial for my English learning. However, I didn't like my teacher's teaching style, so I didn't continue using her recommended mobile learning application, Quizlet. (#2)

4.4.4. Overall appraisal and performance impact in the volatility dimension

Volatility refers to the degree to which individuals are distracted when conducting a task. In this dimension, some factors may effectively sustain learners' focus on a MALL task; others may easily distract learners from a MALL task and impair their overall performance. Most participants stated that if their overall appraisal of and performance in a MALL task was positive, they could remain focused until its completion (e.g., "In the past, while I was doing an English learning task on Kahoot in class, I decided that this was the most effective way to improve my English learning performance, so I tried my best to complete it" #33). Some participants abandoned a MALL task and found other, novel activities when they appraised the MALL task as dull and not improving their learning performance (e.g., "When I was a high school student, I found it so boring to complete repetitive activities on Quizlet for many days, so I started browsing Instagram and Facebook and didn't finish the task assigned by my teacher" #68).

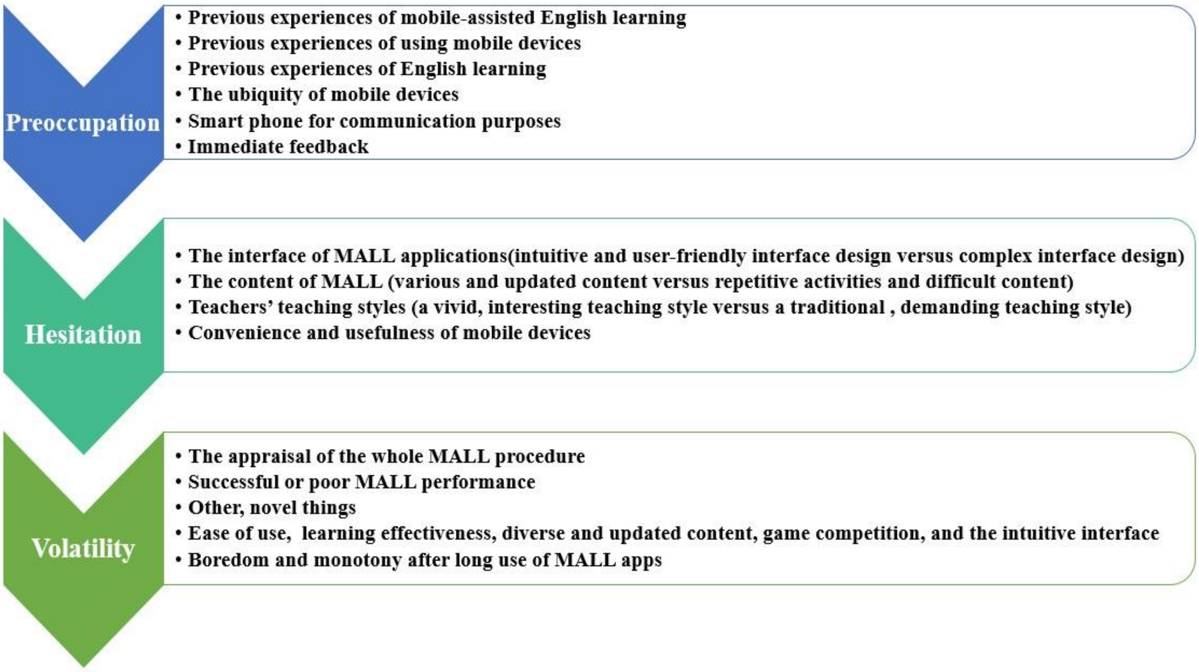
4.4.5. Other novelty interference in the volatility dimension

Most the participants indicated that ease of use, favourable English language learning effectiveness, diverse and updated content, and game competitions could prevent them from being attracted by other, novel things and complete the initiated MALL task (e.g., "I enjoyed quiz games on Socrative because I wanted to beat my classmates. This way I could also enhance my English ability. The activity made me complete the assigned task" #41). Participant 26 noted, "updated content on Kahoot could sustain my interest in learning English on this app to achieve the final goal and complete the task." Participant 13 said, "I didn't have difficulty navigating when I did the Quizlet vocabulary learning task. The user-friendly interface motivated me to complete the task." However, some interviewees mentioned that boredom and monotony after long use of MALL apps might prompt them to find other novelties (e.g., "After working on a MALL task for a long time, I felt a sense of boredom, which made me want to abandon the current task and find other, novel activities to do" #9).

4.5. A summary of factors influencing learner intention changes in the three ACT dimensions

According to results from the survey and interview, we summarised factors influencing learner intention changes in the three ACT dimensions (preoccupation, hesitation, and volatility) in Figure 1. In the preoccupation dimension, previous experiences of mobile-assisted English learning, using mobile devices, and English learning, the ubiquity of mobile devices, smart phone for communication purposes, and immediate feedback may be factors influencing learner intention to disentangle from alternative targets or negative thoughts. In the hesitation dimension, the interface of MALL applications (intuitive and user-friendly interface design versus complex interface design), the content of MALL (various and updated content versus repetitive activities and difficult content), teachers' teaching styles (a vivid, interesting teaching style versus a traditional, demanding teaching style), and convenience and usefulness of mobile devices may influence learner intention to initiate a MALL task. In the volatility dimensions, the appraisal of the whole MALL procedure, successful or poor MALL performance, other, novel things, ease of use, learning effectiveness, diverse and updated content, game competition, the intuitive interface, and boredom and monotony after long use of MALL apps may be factors influencing learners' sustaining intention until a MALL task is completed. We considered the factors in the three ACT dimensions may positively or negatively influence a learner's intention changes from removing negative states, initiating a MALL task, to sustaining their intention until the MALL task is completed.

Figure 1. A summary of factors influencing learner intention changes in the three ACT dimensions



5. Discussion

MALL research in higher education rarely explores factors affecting learners' intention to use MALL from the perspective of psychological traits (Yu, 2020). This study examined factors affecting the behavioural intention to engage in language learning with mobile devices in higher education through the lens of ACT (Kuhl, 1994a). The results of the survey indicated that the three dimensions (preoccupation, hesitation, and volatility) were evident in higher education settings because of their strong, significant correlational relationships. Preoccupation was positively correlated with hesitation, and preoccupation and hesitation were highly correlated with volatility. Each of the three may influence the decision to initiate a MALL task and then sustain focus until its completion.

Previous experiences in three aspects (i.e., MALL, the use of mobile devices, and English learning) may interrupt the focus required for engaging in MALL tasks in the preoccupation dimension (Khany & Amiri, 2018). Positive experiences in these areas help learners disentangle themselves from alternatives and initiate a MALL task because they perceive benefits in MALL, mobile devices, and English language learning. Negative experiences in these areas may result in learners becoming "obsessed" with negative states and feeling afraid of initiating a new MALL task, which conforms to the findings of Khany and Amiri (2018) and Kuhl (1994a).

Several elements (i.e., interface design, content, teaching styles, and convenience and usefulness) may affect hesitation. The results extend the conclusions of previous studies, suggesting that these elements may have a positive or negative influence on hesitation to trigger an action to initiate a MALL task after a decision has been made (Diefendorff et al., 2000; MacIntyre & Blackie, 2012; MacIntyre & Doucette, 2010). On the basis of our findings, we draw the following conclusions. First, an intuitive and user-friendly interface can help learners escape from the vacillation between continuing the initiated MALL task and starting a new activity. Conversely, a complex interface design makes learners consider ending the initiated task and find a new activity to engage in. Second, diverse, interesting MALL content can accelerate the speed of decision making and help learners bypass the hesitation phase to start a MALL task. By contrast, tedious and repetitive MALL content deepens the intensity of hesitation, which in turn encourages learners to engage in a completely new activity. Finally, the perceived ease of use and usefulness of MALL can result in continued MALL use. One possible reason is that learners perceive that little effort is required to complete the MALL task and that their language learning performance has improved, which increases their willingness to continue in the task. If learners do not detect these two benefits, they are likely to abandon the initiated task at the hesitation stage.

In the volatility dimension, learner appraisal of MALL, learning performance, and other novelties may affect their willingness to continue engaging in MALL. For volatile learners, negative appraisal of an initiated MALL task and poor MALL performance discourage them from remaining focused until they complete the task. Conversely, positive appraisal of and performance in MALL play a crucial role in sustaining learner motivation (Harmer, 2007; Khany & Amiri, 2018) until the task is completed. Moreover, when learners perceive that MALL is beneficial, they are less likely to be attracted by other novelties and are more focused on task completion.

5.1. Implications for pedagogical practice

A unique contribution of this study is that it is one of the first attempts to explore the factors that intervene in the intention transformation of university students to initiate a MALL task and continue engaging in the task until it is completed in the three dimensions (preoccupation, hesitation, and volatility) of ACT. The results have pedagogical implications. First, considering positive and negative experiences in three areas (i.e., MALL, mobile device use, and English language learning), instructors may wish to share positive experiences in these areas to help address negative states and accelerate the initiation of a MALL task (Teo et al., 2019).

Second, instructors and developers of MALL applications ought to select intuitive and accessible interfaces and diverse, interesting content. In particular, continually updated content and interactive learning with students from other schools can help students continue attending to an initiated MALL task and avoid their engaging in another task. Third, operation orientation, technical guidance, and immediate assistance need to be offered to students to facilitate their use of MALL applications (Unal & Uzun, 2021).

Fourth, familiarising teachers with mobile technology and MALL applications could alter their teaching beliefs and styles. Mobile-assisted language teaching seminars featuring teachers with experience in MALL can be held. Finally, we suggest that instructors and developers of MALL applications design a variety of small, interesting, and achievable activities to enhance students' sense of learning achievement and remove repetitiveness and predictability from activities. This approach prevents students from turning to novel activities, thus enabling them to concentrate on completing the MALL task.

5.2. Limitations

The study has some limitations. First, the data were collected from several universities in Taiwan. To increase the generalizability of the findings, a similar study would have to be conducted with students of universities in different countries. Second, the study adopted the three variables (preoccupation, hesitation, and volatility) of ACT. Future studies should integrate the variables with other models such as the general extended technology acceptance model for e-learning (GETAMEL, Abdullah & Ward, 2016). Third, the study only focused on the factors influencing learner intention changes in the three ACT dimensions in the context of MALL. However, in future research, other researchers should apply this research and consider carrying out similar studies to investigate what factors may positively and negatively influence the intention transformation process in other contexts (e.g., business marketing, artificial intelligence development, and medicine and society). Finally, a cross-sectional survey on what influences intention transformation in the three dimensions (preoccupation, hesitation, and volatility) was conducted. Future research can explore whether similar effects in the three dimensions of ACT still interfere with the change of student intentions when they are engaged in English language learning on a specific MALL application.

6. Conclusions

Previous experiences in mobile-assisted English learning, mobile device use, and English learning may affect whether university students successfully overcome other thoughts or negative states to initiate a MALL task in the preoccupation dimension. Interface design, content, ease of use, and usefulness of MALL in addition to instructor teaching styles may determine whether a student remains engaged in the initiated MALL task or whether they abandon it to engage in other tasks in the hesitation dimension. The final appraisal, learning performance, and the attraction of other novelties influence whether students remain focused until the initiated MALL task is completed in the volatility dimension. All in all, successful acceptance and completion of MALL tasks depend mainly on student intention (Venkatesh et al., 2003); thus, the determinants affecting intention transformation in the three dimensions of ACT should be investigated to assist students in smoothly completing MALL tasks and to enhance the possibility of mobile device use for language learning in higher education.

The research findings revealed that the factors intervening the three dimensions (preoccupation, hesitation, and volatility) of ACT provide a new research direction for the future MALL and technology-enhanced language learning contexts. Future research is necessarily conducted to obtain deeper understanding of these interference factors or to investigate new interference factors influencing learners' acceptance of the technology for their language learning. In the study, we also developed a scale regarding factors interfering each of the three ACT dimensions in the MALL context to identify what factors might positively and negatively influence learners' intention to accept mobile technology for language learning. Future research can further examine or investigate pedagogical strategies or application function designs to strengthen learners' positive experiences and weaken their negative experiences caused by these interference factors to remove their negative states and other thoughts, and then maintain their learning intention until a MALL task is completed. In addition, many studies have been centered on the effect of mobile technology on vocabulary, listening, speaking, writing, and grammar. However, studies on the interference factors influencing learner intention to accept mobile technology for their language learning through the lens of ACT are still limited.

Conflict of interest

The authors declare that they have no conflicts of interest to this work.

References

Abdullah, F., & Ward, R. (2016). Developing a general extended technology acceptance model for e-learning (GETAMEL) by analysing commonly used external factors. *Computers in Human Behaviour*, *56*, 238-256. https://doi.org/10.1016/j.chb.2015.11.036

Burston, J. (2015). Twenty years of MALL project implementation: A Meta-analysis of learning outcomes. *ReCALL*, *27*(1), 4-20. https://doi.org/10.1017/S0958344014000159

Chang, C. K., & Hsu, C. K. (2011). A Mobile-assisted synchronously collaborative translation-annotation system for English as a Foreign Language (EFL) reading comprehension. *Computer Assisted Language Learning*, *24*(2), 155-180. https://doi.org/10.1080/09588221.2010.536952

Chang, C. C., Liang, C. Y., Yan, C. F., & Tseng, J. S. (2013). The Impact of college students' intrinsic and extrinsic motivation on continuance intention to use English mobile learning systems. *The Asia-Pacific Education Researcher*, *22*(2), 181-192. https://doi.org/10.1007/s40299-012-0011-7

Chen, C. M., & Chung, C. J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, *51*(2), 624-645. https://doi.org/10.1016/j.compedu.2007.06.011

Chen, C.-M., & Hsu, S.-H. (2008). Personalized intelligent mobile learning system for supporting effective English learning. *Educational Technology & Society*, *11*(3), 153-180.

Chen, C. M., Cheng, L. C., & Yang, S. M. (2019). An English vocabulary learning app with self-regulated learning mechanism to improve learning performance and motivation. *Computer Assisted Language Learning*, *32*(3), 237-260. https://doi.org/10.1080/09588221.2018.1485708

Chen, C. P. (2018). Understanding mobile English-learning gaming adopters in self-learning market: The Uses of gratification expectancy. *Computers & Education*, *126*, 217-230. https://doi.org/10.1016/j.compedu.2018.07.015

Chwo, S.-M., Marek, M. W., & Wu, W.-C. (2016). Curriculum integration of MALL in L1/L2 pedagogy: Perspective on research. *Educational Technology & Society*, *19*(2), 340-354.

Crompton, H. (2013). Mobile learning: New approach, new theory. In Z. L. Berge, & L. Y. Muilenburg (Eds.), *Handbook of mobile learning* (pp. 47-57). Routledge.

Crompton, H., & Burke, D. (2018). The Use of mobile learning in higher education: A Systematic review. *Computers & Education*, *123*, 53-64. https://doi.org/10.1016/j.compedu.2018.04.007

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-341. https://doi.org/10.2307/249008

Diefendorff, J. M., Hall, R. J., Lord, R. G., & Strean, M. L. (2000). Action-state orientation: Construct validity of a revised measure and its relationship to work related variables. *Journal of Applied Psychology*, *85*(2), 250−263. https://doi.org/10.1037/0021- 9010.85.2.250

Dörnyei, Z. (2001). *Teaching and researching motivation*. Pearson Education.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative qualitative, and mixed methodologies*. Oxford University Press.

Dörnyei, Z., & Ryan, S. (2015). *The Psychology of the language learner revisited*. Routledge.

Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, *27*(2), 197-216. https://doi.org/10.1017/S0958344014000287

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, *19*(4), 440−452. https://doi.org/10.2307/3151718

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw Hill.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A Review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70-105. https://doi.org/10.1080/09588221.2012.700315

Harmer, J. (2007). *The Practice of English language teaching*. Longman.

Hoi, V. N. (2020). Understanding higher education learners' acceptance and use of mobile devices for language learning: A Rasch-based path modeling approach. *Computers & Education*, *146*, 1-15. https://doi.org/10.1016/j.compedu.2019.103761

Hsieh, J. S. C., Huang, Y. M., & Wu, W. C. V. (2017). Technological acceptance of LINE in flipped EFL oral training. *Computers in Human Behaviour*, *70*, 178-190. https://doi.org/10.1016/j.chb.2016.12.066

Hsu, C. K. (2015). Learning motivation and adaptive video caption filtering for EFL learners using handheld devices. *ReCALL*, *27*(1), 84−103. https://doi.org/10.1017/S0958344014000214

Hsu, C. K., Hwang, G. J., & Chang, C. K. (2013). A Personalized recommendation-based mobile learning approach to improving the reading performance of EFL 118 students. *Computers & Education*, *63*, 327-336.

Hsu, H. T., & Lin, C. C. (2022). Extending the technology acceptance model of college learners' mobile-assisted language learning by incorporating psychological constructs. *British Journal of Educational Technology, 53*, 286-306. https://doi.org/10.1111/bjet.13165

Hwang, G. J., & Fu, Q. K. (2019). Trends in the research design and application of mobile language learning: A Review of 2007-2016 publications in selected SSCI journals. *Interactive Learning Environments*, *27*(4), 567-581. https://doi.org/10.1080/10494820.2018.1486861

Hwang, P. S. (2018). *The Impact of green recruitment activities on the attractiveness of organizational talents: The Perception of personal organizational fit as the intermediary variable and the personal environmental attitude as the interfering variable* [Unpublished master's thesis]. Tung Hai University, Taiwan.

Hwang, W. Y., & Chen, H. S. L. (2013). Users' familiar situational contexts facilitate the practice of EFL in elementary schools with mobile devices. *Computer Assisted Language Learning*, *26*(2), 101-125. https://doi.org/10.1080/09588221.2011.639783

Hwang, W. Y., Huang, Y. M., Shadiev, R., Wu, S. Y., & Chen, S. L. (2014). Effects of using mobile devices on English listening diversity and speaking for EFL elementary students. *Australasian Journal of Education Technology*, *30*(5), 503-516. https://doi.org/10.14742/ajet.237

Jaramillo, F., Locander, W. B., Spector, P. E., & Harris, E. G. (2007). Getting the job done: The Moderating role of initiative on the relationship between intrinsic motivation and adaptive selling. *Journal of Personal Selling & Sales Management*, *27*(1), 59−74. https://doi.org/10.2753/PSS0885-3134270104

Jiang, D., & Zhang, L. J. (2020). Collaborating with "familiar" strangers in mobile-assisted environments: The effect of socializing activities on learning EFL writing. *Computers & Education, 150*, 103841. https://doi.org/10.1016/j.compedu.2020.103841

Kaiser, H. F. (1974). An Index of factorial simplicity. *Psychometrica*, *39*, 31-36. https://doi.org/10.1007/BF02291575

Ke, F., & Hsu, Y. (2015). Mobile augmented-reality artefact creation as a component of mobile computer-supported collaborative learning. *The Internet and Higher Education*, *26*, 33-41. https://doi.org/10.1016/j.iheduc.2015.04.003

Khany, R., & Amiri, M. (2018). Action control, L2 motivational self system, and motivated learning behaviour in a foreign language learning context. *European Journal of Psychology Education*, *33*(2), 337-353. https://doi.org/10.1007/s10212-016-0325-6

Kim, H. S. (2011). Effects of SMS text messaging on vocabulary learning. *Multimedia Assisted Language Learning*, *14*(2), 159-180.

Kim, H. J., Lee, J. M., & Rha, J. Y. (2017). Understanding the role of user resistance on mobile 29 learning usage among university students. *Computers & Education*, *113*, 108-118. https://doi.org/10.1016/j.compedu.2017.05.015

Kuhl, J. (1994a). A Theory of action and state orientations. In J. Kuhl, & J. Beckmann (Eds.), *Volition and personality: Action versus state orientation* (pp. 9−45). Hogrefe & Huber.

Kuhl, J. (1994b). Action vs. state orientation: Psychometric properties of the Action Control Scale (ACS-90). In J. Kuhl, & J. Beckmann (Eds.), *Volition and personality* (pp. 47-59). Hogrefe & Huber Publishers.

Kukulska-Hulme, A., & Shield, L. (2008). Overview of mobile assisted language learning: Can mobile devices support collaborative practice in speaking and listening? *ReCALL*, *20*(3), 271-289. https://doi.org/10.1017/S0958344008000335

Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile learning revolution: Implications for language pedagogy. In C. A. Chapelle, & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 217-233*)*. John Wiley & Sons.

Kvale, S. (1996). *Interviews: An Introduction to qualitative research interviewing*. Sage.

Kvale, S. (2003). The Psychoanalytic interview as inspiration for qualitative research. In P. M. Camic, J. E. Rhodes, & L. Yardley (Eds.), *Qualitative Research in Psychology* (pp. 275-297). American Psychological Association.

Lai, C., & Zheng, D. (2018). Self-directed use of mobile devices for language learning beyond the classroom. *ReCALL, 30*(3), 299-318. https://doi.org/10.1017/S0958344017000258

Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language, Learning and Technolog*y, *17*(3), 135-156.

Lin, C. C., Lin, V., Liu, G. Z., Kou, X. J., Kulikova, A., & Lin, W. L. (2020). Mobile-assisted reading development: A Review from the activity theory perspective. *Computer Assisted Language Learning, 33*(8), 833-864.

Lin, J. J., & Lin. H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A Systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878-919. https://doi.org/10.1080/09588221.2018.1541359

Loewen, S., Crowther, D., Isbell, D., Kim, K., Maloney, J., Miller, Z., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL, 31*(3), 293-311. https://doi.org/10.1017/S0958344019000065

MacIntyre, P. D., & Blackie, R. A. (2012). Action control, motivated strategies, and integrative motivation as predictors of language learning affect and the intention to continue learning French. *System, 40*(4), 533-543. https://doi.org/10.1016/j.system.2012.10.014

MacIntyre, P. D., & Doucette., J. (2010). Willingness to communicate and action control. *System, 38*(2), 161-171. https://doi.org/10.1016/j.system.2009.12.013

MacIntyre, P. D., MacMaster, K., & Baker, S. (2001). The Convergence of multiple models of motivation for second language learning: Gardner, Pintrich, Kuhl and McCroskey. In Z. Dörnyei, & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 461-492). Second Language Teaching & Curriculum Centre.

Nguyen, T.-H., Hwang, W.-Y., Pham, S.-L., & Ma, Z.-H. (2018). User-oriented EFL speaking through application and exercise: Instant speech translation and shadowing in authentic context. *Educational Technology & Society, 21*(4), 129-142.

Nie, J., Zheng, C., Zeng, P., Zhou, B., Lei, L., & Wang, P. (2020). Using the theory of planned behaviour and the role of social image to understand mobile English learning check-in behaviour. *Computers & Education, 156*, 1-13. https://doi.org/10.1016/j.compedu.2020.103942

Nikou, S. A., & Economides, A. A. (2017). Mobile-based assessment: Integrating acceptance and motivational factors into a combined model of self-determination theory and technology acceptance. *Computers in Human Behavior, 68*, 83-95.

Ono, Y., Ishihara, M., & Yamashiro, M. (2015). Blended instruction utilizing mobile tools in English teaching at colleges of technology. *Electrical Engineering in Japan*, *192*(2), 1-11. https://doi.org/10.1002/eej.22717

Park, S. Y., Nam, M. W., & Cha, S. B. (2012). University students' behavioural intention to use mobile learning: Evaluating the technology acceptance model. *British Journal of Educational Technology, 43*(4), 592-605. https://doi.org/10.1111/j.1467-8535.2011.01229.x

Pegrum, M. (2014). *Mobile learning: Languages, literacies, and cultures*. Palgrave Macmillan.

Petersen, K., & Sachs, R. (2016). The Language classroom in the age of networked learning. In R. P., Leow, L. Cerezo, & M. Baralt (Eds.), *A psycholinguistic approach to technology and language learning* (pp. 3-22). De Gruyter.

Pew. (2021). *Mobile fact sheet*. http://www.pewinternet.org/fact-sheet/mobile/

Poushter, J. (2016). Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Centre, 22*(1), 1-44.

Reinders, H., & Benson, P. (2017). Research agenda: Language learning beyond the classroom. *Language Teaching, 50*(4), 561-578. https://doi.org/10.1017/S0261444817000192

Reinders, H., & Pegrum, M. (2017). Supporting language learning on the move. In B. Tomlinson (Ed.), *SLA Research and materials development for language learning* (pp. 219-232). Routledge.

Shadiev, R., W. Hwang, Y., & Huang, Y. M. (2017). Review of research on mobile language learning in authentic environments. *Computer Assisted Language Learning, 30*(3-4), 284-303. https://doi.org/10.1080/09588221.2017.1308383

Stockwell, G. (2007). Vocabulary on the move: Investigating an intelligent mobile phone-based vocabulary tutor. *Computer Assisted Language Learning*, *20*(4), 365-383. https://doi.org/10.1080/09588220701745817

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning & Technology, 14*(2), 95-110.

Teo, T., Zhou, M., Fan, A. C. W., & F. Huang, F. (2019). Factors that influence university students' intention to use Moodle: A Study in Macau. *Educational Technology Research and Development, 67*(3), 749-766. https://doi.org/10.1007/s11423-019-09650-x

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning, 21*(3), 217-228. https://doi.org/10.1111/j.1365-2729.2005.00129.x

Unal, E., & Uzun, A M. (2021). Understanding university students' behavioural intention to use Edmodo through the lens of an extended technology acceptance model. *British Journal of Educational Technology, 52*(2), 619-637. https://doi.org/10.1111/bjet.13046

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478. https://doi.org/10.2307/30036540

Wang, S., & Smith, S. (2013). Reading and grammar learning through mobile phones. *Language Learning & Technology, 17*(3), 117-134.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Sage.

Wu, T.-T., & Huang, Y.-M. (2017). A Mobile game-based English vocabulary practice system based on portfolio analysis. *Educational Technology & Society, 20*(2), 265-277.

Yu, Z. (2020). Extending the learning technology acceptance model of WeChat by adding new psychological constructs. *Journal of Educational Computing Research, 58*(6), 1121-1143. https://doi.org/10.1177/0735633120923772

Li, S. C., & Petersen, K. B. (2022). Does ICT Matter? Unfolding the Complex Multilevel Structural Relationship between Technology Use and Academic Achievements in PISA 2015. *Educational Technology & Society*, *25* (4), 43-55.

Does ICT Matter? Unfolding the Complex Multilevel Structural Relationship between Technology Use and Academic Achievements in PISA 2015

Sandy C. Li[1*] and Karen B. Petersen[2]
[1]Hong Kong Baptist University, HKSAR, PRC // [2]Aarhus University, Denmark // sandyli@hkbu.edu.hk // kp@edu.au.dk
[*]Corresponding author

(Submitted September 21, 2021; Revised December 7, 2021; Accepted February 3, 2022)

ABSTRACT: While infusion of technology into schools has been one of the top priorities of the education reform agenda across the world, findings from many large-scale international assessments indicate that students' use of information and communication technology (ICT) has mixed effects on their academic achievements. In this paper, we argue that these ambivalent findings were due to the oversight of the indirect effects of ICT use mediated by other ICT-related variables. We employed multilevel structural equation modelling to unfold the relationship between students' ICT use and their academic achievements based on PISA 2015 data. The results indicated that students' autonomy in ICT use and students' interest in ICT use were found to have significant positive direct effects on students' academic achievements at both within-school and between-school levels. These two variables played a significant role in mediating the indirect effects of ICT use outside school for schoolwork and ICT resources on students' academic achievements. On the contrary, ICT resources and ICT use at school exerted either no direct effect or a negative direct effect on students' academic achievements and students' perceived autonomy related to ICT use, suggesting that mere provision and use of ICT resources in school did not necessarily guarantee success in student performance. At the school level, school's transformational leadership and collaborative climate helped promote students' autonomy in ICT use.

Keywords: ICT use, Academic achievement, Multilevel analysis, Structural equation modelling, PISA 2015

1. Introduction

Students of today's generation have been immersed in tablets, smartphones and different forms of digital media from a very early age. The rapid advent of information and communication technologies (ICT) in the past two decades has disrupted the world in the ways that people live and interact with one another. ICT is so pervasive that it has become an integral part of our daily-life and one of the key vehicles for driving economic development and leveraging academic attainments. Harnessing new technologies are considered pivotal to education quality, equality, inclusion and life-long learning for all. It is anticipated that the use of ICT helps connect learning across formal and informal contexts in a seamless way (Cai et al., 2019). As such, this global phenomenon has spawned a proliferation of research studies on ICT implementation in schools in the past two decades (Bernacki et al., 2020; Sanders & George, 2017). Nonetheless, the impact of ICT use on students' academic achievements has been equivocal. Large-scale studies on technology integration and student achievements, such as SITES, TIMSS, PISA and PIRLS, have yielded ambivalent results (Bulut & Cutumisu, 2018). The scoping literature review of research articles on PISA assessment published over the past 10 years conducted by Odell et al. (2020) revealed that the relationship between ICT use and academic achievements is ambivalent, and that the relationship varies across subjects, countries and the types of ICT use.

As pointed out by Park and Weng (2020), the statistical methods employed in many large-scale international assessment studies, such as PISA, were mainly hierarchical linear modelling (HLM). To analyze the relationship between the predictors and the multivariate outcome variables such as the mathematics, reading and science scores, a series of HLM analyses have to be conducted separately, which can inflate Type I error. Furthermore, although the abovementioned studies adopted a multi-level approach to data analysis, many of them emphasized on examining how country-level ICT factors, such as GDP per capita and National ICT development index (Hu et al., 2018; Odell et al., 2020; Park & Weng, 2020), impacted on academic achievements without considering the effects of school-level factors, such as school leadership and school climate, which are deemed to be conducive to ICT implementation in school. In addition, at the student-level, none of the above studies examined how the mediation among ICT-related variables actually affected academic achievements. For instance, the provision of ICT resources may not have a direct effect on academic achievements, but it can have a direct effect on promoting students' attitude and autonomy in ICT use, which in turn may have a positive impact on student learning. Yet, HLM is not adequate enough for unfolding these mediation and indirect effects. As such, the inconsistency in results on the relationship between ICT-related variables and academic achievements can be attributed to the methodological inadequacy in data analysis, and the lack of an ecological perspective in examining how the interplay of ICT-related variables and school-level variables actually affected students' academic achievements.

To shed more light on the ongoing controversy over the effect of students' ICT use on their academic achievements, it is necessary to develop an analytical framework that can delineate the complex interplay of various ICT-related variables in mediating the relationship between ICT use and academic achievements within and across different levels.

2. Conceptual framework

ICT implementation in school has long been seen as a complex process. According to the ecological perspective of ICT implementation suggested by Wong and Li (2011), and Li and Choi (2014), the impact of ICT on student learning and academic achievements hinges on a wide spectrum of variables such as student-level, teacher-level and school-level variables. These include students' ICT access, students' competency in ICT use, students' attitude towards ICT use, students' autonomy in ICT use, teacher's collaboration and school leadership. The subtle within-level and cross-level interactions among these variables play a significant role in mediating the relationship between ICT use and students' academic achievements. Ignoring the ecological dynamics of these variables in ICT implementation might lead to significant discrepancies in the results. This is seen as one of the possible reasons accounting for the ambivalent findings regarding the relationship between ICT use and students' academic achievements found in many large-scale international assessment studies (Park & Weng, 2020).

In the context of large-scale international assessments such as PISA, it remains unknown about the within- and between-level interactions among ICT-related variables mediate the relationship of ICT use and students' academic achievements. To bridge these gaps, the purpose of the present study was to unfold the complex relationship between ICT use and students' academic achievements by constructing multilevel structural equation models in which variables can interact with one another along multiple paths and across different levels. Guided by the ecological perspective of ICT implementation mentioned above, we specifically examined (1) the direct effects of various ICT-related variables on academic achievements at the student and school levels, (2) the mediation of students' perceived autonomy and interest in ICT use acting between students' ICT use and their academic achievements, and (3) the role of school-level variables such as, school leadership and school's collaborative climate in shaping the impact of ICT use on academic achievements.

2.1. ICT use and academic achievements

There has been a long debate on the relationship between students' ICT use and their academic achievements. From the constructivist perspectives, students' ICT use facilitates the development of learner autonomy, expression of thoughts and negotiation of meaning among learners, which eventually lead to enhancement in student learning outcomes (Chiao & Chiu, 2018). Nevertheless, results from large-scale international assessment on ICT use and academic achievements are inconclusive. Zhang and Liu (2016) examined the trends of relationships between ICT use and students' academic achievements across the five waves of PISA studies from 2000 to 2012, and identified that students' ICT use was negatively correlated with their science and mathematics achievements.

In PISA studies, students' ICT use can be categorized into ICT use at school, ICT use outside school for schoolwork, ICT use for entertainment and ICT use for social interaction (OECD, 2016b). Different types of students' ICT use manifested different relationships with students' academic achievements across the academic subjects being examined (Odell et al., 2020). Based on PISA 2015, Hu et al. (2018) and Gómez-Fernández and Mediavilla (2021) revealed that students' ICT use for entertainment correlated positively with their academic achievements while negative correlations were found between students' ICT use outside school for schoolwork and achievements. On the other hand, by analyzing the Dutch PISA 2015 sample, Gubbels et al. (2020) found that students with moderate ICT use outside school for schoolwork had the highest performance in reading, whereas frequent use of ICT outside school for leisure correlated negatively with reading scores. Students' ICT use for social interaction and ICT use for entertainment were found to correlate negatively with students' academic achievements (Gómez-Fernández & Mediavilla, 2021; Hu et al., 2018; Odell et al., 2020; Park & Weng, 2020). Likewise, negative correlation between ICT use at school and academic achievements was also identified (Hu et al., 2018).

While the relationships between students' ICT use and their academic achievements were inconclusive, findings from some studies indicating that students' ICT use for learning, social interaction or leisure helped promote students' autonomy and interest in harnessing technology (Burbat, 2016; Honarzad & Rassaei, 2019; Liu et al., 2018). In short, if students are given ample opportunities to harness technology through school work and daily activities, they may develop the ability to use technology to support their own learning during private time. Thus, students' ICT use may exert an indirect effect on academic achievements via students' autonomy and interest in ICT use. However, little is known about these mediation effects in large-scale assessments such as PISA. Thus, the purpose of the present study was to unfold how the interplay of these variables affected academic achievements.

2.2. Autonomy in ICT use, interest in ICT use, competency in ICT use and academic achievements

In the context of PISA studies, students' autonomy in ICT use, interest in ICT use and competency in ICT use refer respectively to students taking control of their learning through ICT use, students' intrinsic motivation to use ICT, and students' ICT-related skills and knowledge (Park & Weng, 2020).

In the trend analysis of five waves of PISA studies from 2000 to 2012 conducted by Zhang and Liu (2016), students' competency in ICT use was found to correlate positively with science and mathematics and reading achievements. Positive associations between students' autonomy in ICT use and academic achievements were identified in PISA 2015 (Gómez-Fernández & Mediavilla, 2021; Hu et al., 2018; Park & Weng, 2020; Petko et al., 2017). Similar relationships between students' interest in ICT use and academic achievements were also identified in the above studies. A non-linear relationship was found between students' interest in ICT use and reading achievements in the Dutch PISA 2015 sample (Gubbels et al., 2020). It was found that students with moderate interest in ICT use had the highest scores in reading. Inconsistent results of the relationship between students' ICT interest and academic achievements were also found across countries. In the regression analysis of PISA 2015 data conducted by Meng et al. (2019), students' interest in ICT use was found to correlate positively with students' academic achievements in China, in contrast to the negative correlation in Germany.

As illustrated above, these ICT traits exert significant impacts on students' academic achievements and can play a pivotal role in mediating the influence of other ICT-related variables, such as ICT use and ICT availability on academic achievements.

2.3. ICT availability at school, ICT availability at home and academic achievements

In PISA studies, ICT availability can be categorized into ICT availability at school and at home. Building the ICT infrastructure and providing students the access to ICT resources is often seen as instrumental to ICT implementation in school (Hu et al., 2018). Zhang and Liu (2016), after controlling for demographic variables at the school level, found that the number of Internet-connected computers available to students exert positive influence on academic achievements across the five waves of PISA studies. Results also show that higher availability of computers per student in the schools correlated positively with students' academic achievements (Gómez-Fernández & Mediavilla, 2021). Similarly, ICT availability at home was found to correlate positively with students' reading scores in PISA 2015 by the analysis conducted by Yalcin (2018), in contrast to the negative correlation identified by Hu et al. (2018). Non-linear relationship was also revealed in the relationship between ICT availability at school and at home, and students' reading scores (Gubbels et al., 2020).

Similar to students' ICT use, ICT availability at school and at home may exert an indirect effect on students' academic achievements via students' autonomy in ICT use and students' interest in ICT use. These mediation or indirect effects remain underexplored in PISA studies.

2.4. School's transformational leadership, teachers' collaborative climate and academic achievements

In comparison to other system-wide education reform initiatives, infusing ICT in schools is often seen as more intricate and challenging. Effective implementation hinges not only on the provision of physical resources, it also depends on an array of organizational factors such as school leadership, the collaborative climate within a school (Amghar, 2019; Szeto, 2020). Among various school-level factors, fostering communities of practice for teachers, provision of a participative governance structure that empower teachers to make autonomous decisions (Avidov-Ungar & Hanin-Itzak, 2019), and principals' leadership, particularly, transformational leadership was found to be pivotal to sustaining change in school (Wu et al., 2020). Transformational leadership is often conceptualized as a leadership trait in which school leaders possess the capacity to (1) establish common goals and shared visions, (2) cultivate mutual trust and support among staff, (3) empower teachers to take risk and experiment with new practices, and (4) support professional development (Bush, 2018;). Demir (2021) argued that factors affecting teachers' adoption of new practices can be irrational or sociocultural, and that trust, collegiality, social support and professional exchange are instrumental social forces that help sustain change and innovations in school. Tam et al. (2018) identified these social forces as the essential social fabrics of a school which enable teachers to have access to expertise and collegial support, and make them feel safe in risk-taking. Transformational leadership played a critical role in shaping teachers' collaborative climate for change and exerted an indirect effect on fostering innovative practices and students' academic achievements (Li & Choi, 2014). Small positive associations between teachers' collaborative climate and students' academic achievements were identified in the German PISA 2012 sample (Mora-Ruano et al., 2019). On the other hand, working on the PISA 2015 data, Wu et al. (2020) unveiled that school leadership had a direct positive relationship with teachers' collaborative climate and students' science achievements. In our present study, we attempted to examine how the interplay of school leadership, teachers' collaborative climate and ICT-related variables such students' autonomy in ICT use and students' interest in ICT use impacted on students' academic achievements at the school level.

2.5. Socioeconomic status and academic achievements

Socioeconomic status (ESCS) in PISA studies is operationally defined as a measure of students' access to family resources including human social, cultural, and financial capitals, which identify the social position of the student's family (Avvisati, 2020). ESCS was found to have a strong correlation with students' academic achievements (Chiao & Chiu, 2018). In the present study, in order to control for the influence of student's socioeconomic status on academic achievements, ESCS was used as a control variable for the analysis at the student-level and school-level.

2.6. The present study

Apparently, the analytical methods employed in most of studies on large-scale international assessments are mainly regression analysis or hierarchical linear modelling (HLM) which is inadequate to unfold the complex relationships between ICT-related variables and students' academic achievements, particularly, the subtle indirect effects exerted among variables. To circumvent these limitations, we employed multi-level structural equation modelling (MSEM) techniques to examine how ICT-related variables compete with one another and how they mediate the impact of ICT on students' academic achievements. MSEM enables variables at different levels to interact with one another along multiple paths.

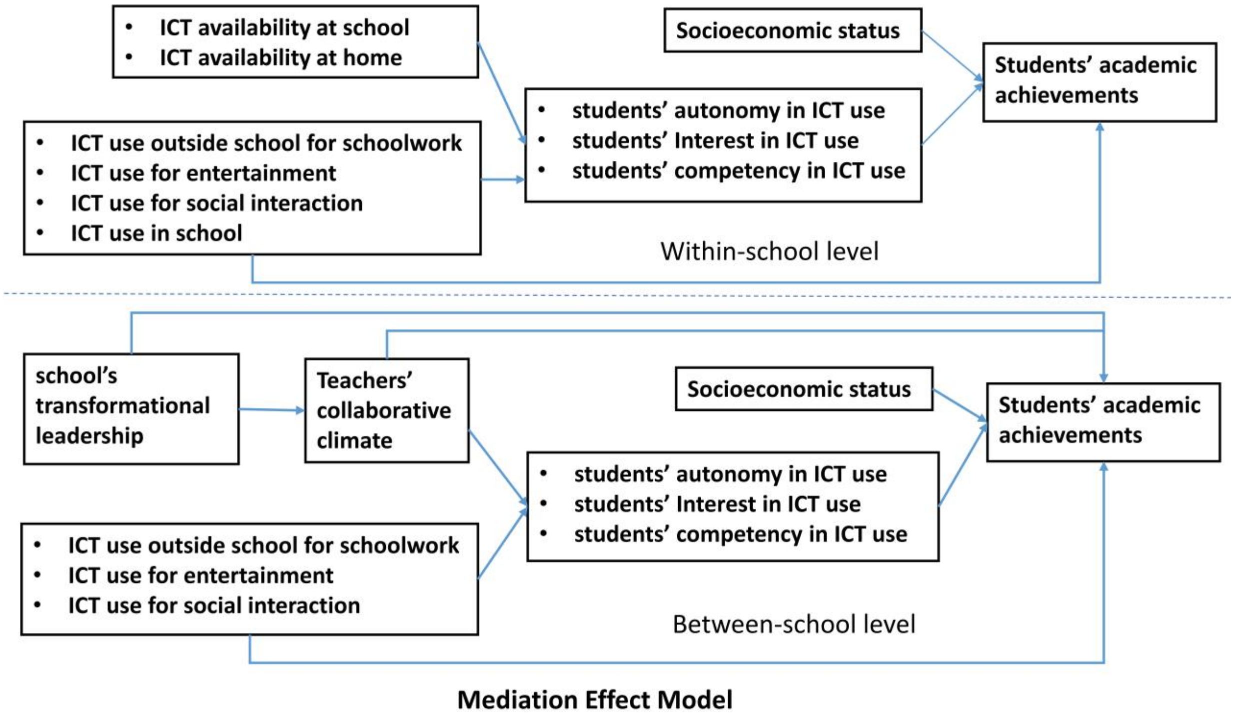
We hypothesized that students' autonomy and interest in ICT use played a mediating role in the relationship between students' ICT use and students' academic achievements. In short, students' ICT use could help develop their autonomy and interest in ICT use which subsequently influenced on students' academic achievements. At the school level, as discussed in previous section, school's transformational leadership had a positive impact on teachers' collaborative climate while school's collaborative climate was found to be conducive to effective ICT implementation. Thus, we anticipated that teachers' collaborative climate would have a positive impact on students' ICT-related latent traits, such as students' autonomy, interest and competence in ICT use. As such, we would like to examine how teachers' collaborative climate mediated the relationship between school's transformational leadership and students' ICT-related traits. The conceptual framework depicted in Figure 1 highlights the complex structural relationship among students' academic achievements, ICT use, ICT related traits and various contextual variables at both the within-school and the between-school levels.

Specifically, two-level random intercept structural equation models were constructed based on the PISA 2015 dataset. Student-level variables included students' academic achievements, students' autonomy in ICT use, students' interest in ICT use, and students' competence in ICT use, while school-level variables encompassed school's transformational leadership, teachers' collaborative climate and school-level ICT resources. Based on this framework, we examined if the multi-level contextual variables had any direct and/or indirect effects in mediating the impact of ICT use on student achievements. The research questions are threefold:

- RQ1. Based on the PISA 2015 dataset, to what extent did students' autonomy in ICT use, students' interest in ICT use, students' competence in ICT use, and their ICT use in various contexts impact on students' academic achievements?
- RQ2. To what extent did students' autonomy in ICT use and students' interest in ICT use mediate the impact of students' ICT use and ICT availability on students' academic achievements?

- RQ3. Did school's transformational leadership and teachers' collaborative climate have a role to play in shaping the impact of ICT use on students' academic achievements?

Figure 1. A conceptual framework depicting a multilevel structural relationships among students' academic achievements, ICT use and various ICT-related variables



3. Method

3.1. Data source and sample

The presented study was grounded on the data derived from the large-scale comparative study: Programme for International Student Assessment (PISA) conducted in 2015, under the auspices of Organization for Economic Co-operation and Development (OECD). Administered in every three years, PISA assesses how well 15-year-old students apply their knowledge and skills in three key domains: reading, mathematics and science (with a stronger focus on a selected domain in each three-year cycle) to address challenges that emerge from their everyday-life experiences. In PISA 2015, a stronger focus was centered on science (OECD, 2016b).

The main reason for choosing PISA 2015 dataset is that it contains variables such as ICT use outside school for schoolwork and teachers' perceived transformational leadership which were not included in the previous cycles of PISA and PISA 2018 studies.

We utilized the school, teacher and student questionnaire of PISA 2015 dataset to examine if students' ICT use in various contexts had any impact on their academic achievements among those high-achieving countries and economies participated in PISA 2015, and how this impact was mediated by students' autonomy in ICT use and students' interest in ICT use.

We initially selected those participating countries or economies whose national mean science literacy scores were ranked top 20 in PISA 2015. Among these 20 countries or economies, only 7 of them participated in all the four questionnaire surveys relevant to the present study: Cognitive test scores, Student Questionnaire, ICT familiarity Questionnaire, and Teacher Questionnaire. These 7 countries or economies, which include Chinese Taipei, Macau, Hong Kong, Beijing-Shanghai-Jiangsu-Guangdong (BSJG China), Korea, Australia and Germany, were finally selected for our study. This selected sample comprised altogether 1,847 schools, 9024 teachers and 53,999 students.

3.2. Data analysis

3.2.1. Variables

Variables for the present study were selected from the Cognitive tests, ICT Familiarity Questionnaire, Teacher Questionnaire and Student Questionnaire of PISA 2015. As listed in Table 1, some variables were composite variables created by the PISA research team, while others were latent variables and their associated indicators.

Table 1. List of within-school level and between-school level variables

Student-level (Within-school level)	School-level (Between-school level)	Variable label/ question items in PISA 2015 Dataset
Variables from Cognitive Tests		
ACW	ACB	Students' academic achievements
SCIE	random intercept	Plausible values 1-10 in Science
MATH	random intercept	Plausible values 1-10 in Mathematics
READ	random intercept	Plausible values 1-10 in Reading
Variables from ICT Familiarity Questionnaire		
SOIAICT		Students' ICT use for social interaction
COMPICT		Students' competence in ICT use
USESCH		Use of ICT at school in general
HOMESCH		ICT use outside of school for schoolwork
ENTUSE		ICT use outside of school leisure
ICTSCH		ICT available at School Index
ICTRES		ICT resources at home
ATW	ATB	Students' autonomy in ICT use
IC015Q02NA	random intercept	If I need new software, I install it by myself.
IC015Q03NA	random intercept	I read information about digital devices independently.
IC015Q05NA	random intercept	I use digital devices as I want to use them.
IC015Q07NA	random intercept	If I have a problem with digital devices I start to solve it on my own.
IC015Q09NA	random intercept	If I need a new application, I choose it by myself.
ITW	ITB	Students' interest in ICT use
IC013Q04NA	random intercept	The Internet is a great resource for obtaining information I am interested in (e.g., news, sports).
IC013Q05NA	random intercept	It is very useful to have social networks on the Internet.
IC013Q11NA	random intercept	I am really excited discovering new digital devices or applications.
IC013Q13NA	random intercept	I like using digital devices.
Variables from Teacher Questionnaire		
-	TCLEAD	Teachers view on school's transformational leadership
-	EXCHT	Teachers' collaborative climate for exchange and co-ordination for teaching
Variables from Student Questionnaire		
ESCS		Index of economic, social and cultural status
W_FSTUWT	-	Final adjusted student weight
-	W_SCHGRNRABWT	Final adjusted school weight

3.2.2. Students' academic achievements

In PISA 2015, science literacy was measured by students' procedural and epistemic knowledge about science. For mathematics, PISA assessed the extent to which students were able to apply mathematics to solve real-world problems. Students' reading literacy was assessed based on how well they comprehended, used and reflected on written texts. As students were assessed with only a small subset of the total item pool, PISA employed imputation methods to report students' academic achievements in order to reduce measurement error. For each literacy domain, PISA used 10 plausible values to represent students' performance, e.g., PV1SCIE – PV10SCIE for science, PV1MATH – PV10MATH for mathematics, and PV1READ - PV10READ for reading. For any analysis involving estimates of students' academic achievements, it should be conducted 10 times (each with one selected plausible value), and the final estimate is obtained by pooling results of the 10 individual analyses. In the present study, we defined students' academic achievements as a latent variable which comprising the math score, science score and reading score as indictors. The within-school level and between-school level components of students' academic achievements were denoted as ACW and ACB respectively (see Table 1).

3.2.3. Student's socioeconomic status

In Student Questionnaire, PISA created an index, ESCS for gauging student's economic, social and cultural status. In the present study, ESCS served as controlled variables. We would like to see how students' ICT use in various contexts, students' autonomy in ICT use and students' interest in ICT use impacted on their academic achievements after controlling the effects due to ESCS.

3.2.4. ICT availability, students' ICT use in various contexts, students' autonomy and interest in ICT use

Based on the ICT Familiarity Questionnaire, PISA created a number of composite variables to measure students' ICT use, for instance, ICT availability at school and at home indices, and ICT resources in general were measured with the composite variables ICTSCH, and ICTRES respectively. According to the PISA 2015 technical report, the composite variable ICTRES comprises six items, asking students about the ICT availability at home including educational software, a link to the Internet, the number of cell phones with Internet access, computers, tablets and e-book readers.

Students' ICT use was differentiated by the location and purpose of technology use, e.g., use of ICT at school in general (USESCH), ICT use outside school for schoolwork (HOMESCH) and ICT use for leisure (ENTUSE). In addition, students' traits related to ICT use included students' competence in ICT use (COMPICT) and social use of ICT (SOIAICT). In addition, students' ICT related traits also included students' interest in ICT use and students' autonomy in ICT use. For multilevel structural equation modelling, the variation of each of these variables were partitioned into a within-school level and a between-school level components for analysis. Taking HOMESCH as an example, the within-school level component of HOMESCH is the variance of individual students' scores about their school mean scores, whereas, the between-school level component of HOMESCH represents the variance of individual school mean scores about the entire sample mean. Thus, the greater the between-school level variance, the greater the diversity is found among schools in terms of their mean scores of HOMESCH.

As illustrated in their technical report, PISA used 5 items: IC015Q02NA, IC015Q03NA, IC015Q05NA, IC015Q07NA and IC015Q09NA to measure students' autonomy in ICT Use. In the present study, the within-school level and between-school level components of students' autonomy in ICT use were denoted as ATW and ATB respectively. To improve model fitting in two-level CFA, IC015Q05NA was deleted from ATB. For students' interest in ICT use, PISA used 6 items to measure students' ICT interest. In the present study, the number of items were reduced to 4 to enhance model-fitting after confirmation factor analysis. The 4 items were: C013Q04NA, IC013Q05NA, IC013Q11NA, and IC013Q13NA as shown in Table 1. Similarly, for the purpose of multilevel structural equation modelling, the variation of students' interest in ICT use was partitioned into within-school level and between-school level components denoted as ITW and ITB respectively.

3.2.5. School's transformational leadership and teachers' collaborative climate

In Teacher Questionnaire, PISA created a composite variable, TCLEAD to gauge teachers' views on school's transformational leadership which include teachers' involvement in decision-making, principal's awareness of teachers' needs and respect for teachers as professionals, principal's capacity to inspire innovative ideas and build consensus with teachers in priority- and goal-setting. Teachers' collaborative climate was measured by teacher's exchange and coordination for teaching via EXCHT.

3.2.6. Multilevel structural equation modelling (MSEM)

As the structure of PISA data was intrinsically hierarchical, in which students and teachers data were nested in schools. Treating the single level data as independent observations may result in underestimating standard errors of regression coefficients and overstating statistical significance. In the present study, two-level random intercept structural equation models were constructed in which the variations in the variables selected from the student questionnaires, such as, HOMESCH, ENTUSE, SOIAICT, INTICT, COMPICT and ESCS, were partitioned into a within-school level and a between-school level components. While ACW, ATW and ITW were modeled as the within-school level components of students' academic achievements, students' autonomy in ICT use, and students' interest in ICT use respectively, ACB, ATB and ITB are the between-school components representing random intercepts of the models. The between-school level variables, TCLEAD and EXCHT were derived by averaging the aggregated scores of teachers' responses collected from each school. Based on the conceptual framework, a two-level random-intercept structural equation model: Mediation Effect Model was constructed to examine the complex structural relationships among students' academic achievements, students' ICT use in various contexts, students' competency in ICT use, students' autonomy in ICT use and students' interest in ICT use.

4. Results

4.1. Mediation effect model

Based on the conceptual framework, a two-level mediation effect model was constructed (see Figure 2). The fitting indices given in Table 2 indicate that the RMSEA indices of the three models are all smaller than 0.05 and all CFI and TLI are greater than or close to 0.9, showing that the three models are of good fit.

Table 2. Fitting indices of two-level structural equation models

	df	χ^2	RMSEA	CFI	TLI
Mediation effect model	245	5306.467	0.023	0.90	0.89

At the within-school level, as shown in Figure 2, ATW, ITW, and HOMESCH exerted a positive direct effect on students' academic achievements with a loading of 0.309, 0.263 and 0.046 respectively, in contrast to the negative effects exerted by ENTUSE, SOIAICT, COMPICT and USESCH which ranged from -0.147 to -0.163. However, as ENTUSE, SOIAICT, COMPICT exerted a moderate to large positive direct effect on ATW and ITW, ranging from 0.109 to 0.452, they had a positive indirect effect on students' academic achievements mediated by ATW and ITW.

ICTRES exerted a large direct effect on ITW and COMPICT, and a small effect on ATW.

Summing over the direct and indirect effects, HOMESCH, ICTRES and COMPICT had a positive total effect on students' academic achievements (see Table 3). USESCH and ICTSCH were the two ICT-related variables showing no effect or a negative effect on ATW, ITW and students' academic achievements.

Table 3. Total effects and indirect effects exerted by ICT-related variables and school-level variables on students' academic achievements

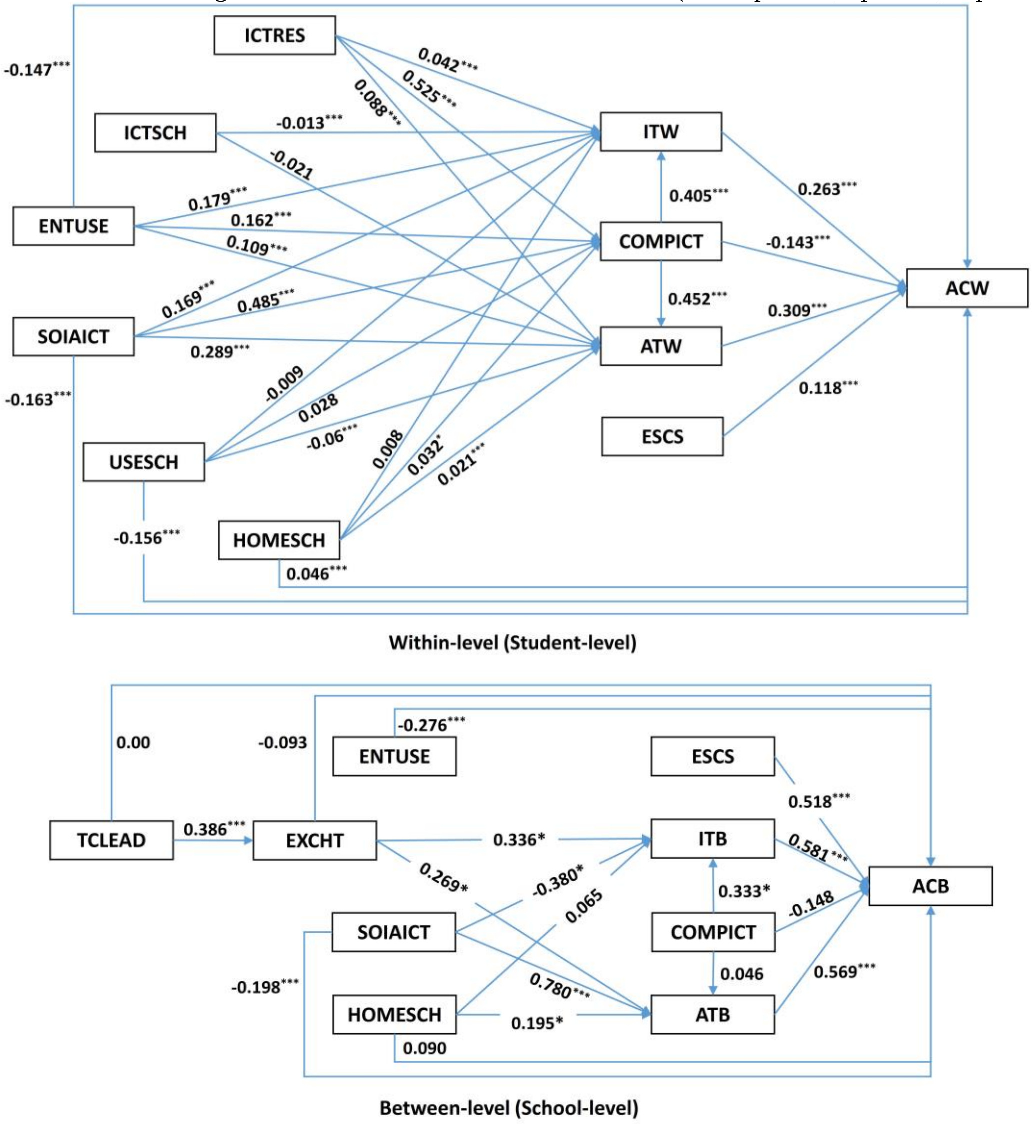
	ACW		ACB	
	Total effect	Total indirect effect	Total effect	Total indirect effect
ATW	0.305^{***}	-		
ATB	-	-	0.529^{**}	-
ITW	0.260^{***}	-	-	-
ITB	-	-	0.584^{**}	-
COMPICT	0.102^{***}	0.244^{***}	0.073	0.225
SOIAICT	0.020	0.179^{***}	0.022	0.199
ENTUSE	-0.049^{***}	0.096^{***}	-0.272	-
HOMESCH	0.057^{**}	0.012^{*}	0.240^{**}	0.150^{*}
USESCH	-0.174^{***}	-0.018^{***}	-	-
ICTRES	0.038^{***}	0.038^{***}	-	-
ICTSCH	-0.010^{*}	-0.010^{*}	-	-
EXCHT	-	-	0.276^{**}	0.371^{***}
TCLEAD	-	-	0.107^{*}	0.107^{*}

Note. (W) Within-school level; (B): Between-school level; $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

At the between-school level, ATB and ITB exerted a large direct effects on students' academic achievements with a loading of 0.569 and 0.518 respectively. HOMESCH exerted a moderate positive effect on ATW. As a result, HOMESCH had a positive indirect effect on students' academic achievements with a loading of 0.057 mediated by ATW (see Table 3). On the other hand, EXCHT exerted a strong positive direct effect on ATB and ITB with a loading of 0.269 and 0.336 respectively (see Figure 2). As a result, EXCHT manifested moderate positive total effects on students' academic achievements with the indirect effects mediated by ATB and ITB.

TCLEAD exerted a strong positive direct effect on EXCHT (0.393), resulting in a small positive total effect (0.107) on students' academic achievements.

Figure 2. Factor loadings of the two-level mediation effect model (*Note.* $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$)



5. Discussion and conclusion

The results of the present study reveal that students' autonomy in ICT use and students' interest in ICT use at both the within-school and between-school levels, exerted a significantly large positive direct effect on students' academic achievements across students' academic achievements comprising science, mathematics and reading after controlling for the effects of ESCS. In addition, these latent traits also played a significant role in mediating the impact of ICT related variables on students' academic achievements. These ICT related variables include students' ICT use outside school for schoolwork, students' ICT use for leisure (ENTUSE) and students' competence in ICT use (COMPICT) and students' ICT use for social interaction (SOIAICT).

The results of the present study corroborate with findings from studies on student autonomy, indicating that students' perceptions of autonomy and academic competence predict students' learning engagement (Hafen et al., 2012). As explicated by Skinner et al. (2008), students tend to look for and flourish in environments where they are empowered to exercise their autonomy and apply their knowledge. In short, enhancing students' autonomy may lead to an increase in their engagement in learning. González and Paoloni (2015) echoed that students' autonomy in ICT use positively predicted their motivation, metacognitive strategies and learning performance. By taking a close look at students' autonomy in ICT use as defined in PISA 2015, it is not difficult to see that the items were devised for assessing students' metacognitive abilities related to ICT use, e.g., "If I have a problem with digital devices I start to solve it on my own," "I read information about digital devices independently," etc. So, this latent trait is not about the freedom students enjoyed in using ICT, but their abilities to harness and use technology for learning. Similarly, students' ICT interest in PISA 2015 is more than their liking for ICT. It probes into students' epistemological beliefs about ICT use and motivation to advance their knowledge related to ICT use, e.g., "The Internet is a great resource for obtaining information I am interested in (e.g., news, sports, dictionary)," "I am really excited discovering new digital devices or applications," etc. In short, the two latent traits seemingly encompass the necessary attributes or competence that enable students to move away from surface learning. Thus, these latent traits related to ICT use apparently played a pivotal role in enhancing students' engagement in learning and academic achievements.

Regarding students' competence in ICT use (COMPICT), while at the within-school level, its direct effect on students' academic achievements were all negative, its indirect effect as mediated by students' autonomy in ICT use was positive and significant, resulting in a positive and significant total effect exerting on students' academic achievements (see Table 2). In a similar fashion, students' ICT use for social interaction (SOIAICT) and for leisure (ENTUSE) exerted a negative effect on students' academic achievements. There were a lot of negative connotations about the impact of students' social use of ICT on their academic achievements. As such, SOIAICT and ENTUSE has been regarded as a negative determinant of academic success and considered a nuisance to student learning (Hu et al., 2018; OECD, 2016a). Yet, taking a closer look at the MSEM results, both SOIAICT and ENTUSE exert a significant positive indirect effect on students' academic achievements as mediated by ATW and ITW at the within-school level and as mediated by ATB and ITB at the between-school level, though their resulting total effects remain negative or insignificant. The effects of students' social use of ICT should deserve more attention in future studies. In particular, it is worthy of examining whether a nonlinear relationship exists between students' academic achievements and their ICT use for leisure and social interaction as this kind of nonlinear relationship was revealed in some PISA studies discussed above (Odell et al., 2020), indicating that moderate or regulated social use of ICT may have a positive impact on students' performance.

It is also noteworthy that, while students' ICT use outside school for schoolwork were found to have either no effects or negative effects on students' academic achievements in a number of studies on PISA data (Hu et al., 2018; Zhang & Liu, 2016), the results of the present study indicate that students' ICT use outside school for schoolwork (HOMESCH) exerts a significant positive direct effect on students' academic achievements at the student level. It is noteworthy that the indirect effects of HOMESCH acting on students' academic achievements are mediated by students' autonomy in ICT use and students' interest in ICT use. This indicates that students' use of ICT outside school for schoolwork possesses the necessary affordance for nurturing students' autonomy and interest in ICT use which leads to their success in academic performance. We argue that informal contexts such as the home environments offer students a more relaxed, secured and autonomous learning space where they can explore, select and orchestrate different technologies for problem-solving, and that this kind of latent ability can be transferred and applied across disciplines. Thus, in future PISA studies, it is worth probing deeper into the connection between the types of schoolwork assigned to students, students' autonomy in ICT use and their generic cognitive and metacognitive skills.

Likewise, ICT availability at home (ICTRES) exerted a positive indirect effect on students' academic achievements mediated by students' autonomy in ICT use and students' interest in ICT use. Obviously, the availability of ICT resources at home would amplify students' capacity in accomplishing their schoolwork and offer more opportunities for students to develop their autonomy in ICT use. Nonetheless, ICT availability at school (ICTSCH) and the general use of ICT at school (USESCH) did not manifest the same effects on students' academic achievements as compared to ICTRES and HOMESCH. Interestingly, USESCH and ICTSCH exerted either no effect or a negative effect on students' autonomy in ICT use and their academic achievements. This suggests that mere provision and use of ICT resources at school did not necessarily guarantee success in student performance. It depends on how technology is being used pedagogically and whether students are able to develop their autonomy in learning with technology in and beyond the classroom processes.

Looking from a broader perspective, teachers' professional exchange and coordination for teaching (EXCHT) exerted significant positive indirect and total effects on Science, Mathematics and Reading which were mediated by students' perceived autonomy related to ICT use (ATB). While EXCHT exerted a significant positive direct effect on ATB, teachers' perceived transformational leadership impacted positively on EXCHT. As mentioned in previous sections, there have been ample studies suggesting that principal's transformational leadership is conducive to cultivating a collaborative climate in which teachers are empowered to experiment with new practices related to pedagogical use of emerging technologies. So, it would be interesting to examine if teachers' pedagogical use of technology mediates the impact of school leadership and school collaborative climate on students' perceived autonomy related to ICT use in future PISA studies. From a methodological point of view, multilevel structural equation modelling helps unravel the complex structural interplay between variables. By teasing out the direct effects as well as indirect effects between variables, one can gain a more complete picture for discerning the impact of ICT use on students' academic achievements.

In sum, the findings of the present study can be summarized as follows:

RQ1. Based on the PISA 2015 dataset, to what extent did students' autonomy in ICT use, students' interest in ICT use and students' competence in ICT use, ICT availability at home and ICT availability at school impact on students' academic achievements?

Based on the results derived from the multilevel structural equation model, students' autonomy in ICT use and students' interest in ICT use were found to the determining variables which exert a large positive effect on students' academic achievements at both the student and school levels. Among students' ICT use in various contexts, students' ICT use outside school for schoolwork was found to exert a positive direct on students' academic achievements. Students' ICT use at school, students' ICT use for social interactions and students' ICT use for leisure were found to exert a negative direct effect on students' academic achievements.

RQ2. To what extent did students' autonomy ICT use and students' interest in ICT use mediate the impact of students' ICT use and ICT availability on students' academic achievements?

Students' autonomy in ICT use and students' interest in ICT use played a pivotal role in mediating the positive effects of ICT-related variables on students' academic achievements. These ICT-related variables include students' ICT use outside school for schoolwork, students' ICT use for social interaction, students' ICT use for entertainment, students' competency in ICT use and ICT availability at home. As a results, students' ICT use outside school for schoolwork, students' competency in ICT and use ICT availability at home manifested a positive total effect on students' academic achievements.

RQ3. Did school's transformational leadership and teachers' collaborative climate have a role to play in shaping the impact of ICT use on students' academic achievements?

At the school level, teachers' collaborative climate exerted a strong direct effect on students' autonomy in ICT use and students' interest in ICT use. As a result, it had a positive total effect on students' academic achievements, with a positive indirect effect mediated by students' autonomy in ICT use and students' interest in ICT use. Likewise, school's transformational leadership exerted a positive direct effect on teachers' collaborative climate, resulting in a positive total effect on students' academic achievements, with an indirect effect mediated by teachers' collaborative climate, students' autonomy in ICT use and students' interest in ICT use.

6. Limitations, implications and future directions

One of the limitations of this study is that among the top 20 high-achieving countries or economies, only 7 of them participated in all the four questionnaire surveys relevant to the present study. As such, 7 countries or economies were involved in the present study. Nonetheless, this selected sample comprised altogether 1,847 schools, 9024 teachers and 53,999 students. To further deepen our understanding of the impacts of ICT use across different regions, selecting data from a larger sample of countries is necessary. Nonetheless, the purpose of this study is not to make any over-generalized claims, but to gain more insights into what and how ICT related factors impinged on students' success in high-achieving countries and economies. Thus, we hope the findings of this study could shed lights on discerning the impact of ICT use on students' academic achievements and pave the way for further studies.

The pedagogical implication of this study is that empowering students to learn with technology is seemingly the key for leveraging the potential of ICT in education. Students need to develop their affection and sense of autonomy or ownership in using ICT to support their daily work. The mere provision of technology at school is not adequate to promote deep learning. In and out of the classroom, it is thus necessary to provide students more opportunities to engage in meaningful learning with the support of technology.

Nonetheless, further research is needed to explore (1) how students' autonomy in ICT use is associated with students' cognitive and metacognitive abilities; (2) the nonlinear effect of students' social use of ICT on students' academic achievements; (3) how pedagogical factors come into play in students' academic achievements; and (4) how teachers' pedagogical use of technology engenders students' autonomy in ICT use and mediates the relationship between school collaborative climate and students' autonomy in ICT use. These are a few possible future research directions which should deserve more attention.

Reference

Amghar, A. (2019). School-based leadership in the education reform agenda in Morocco: An Analysis of policy and context. *International Journal of Leadership in Education*, *22*(1), 102-115. https://doi.org/10.1080/13603124.2018.1543538

Avidov-Ungar, O., & Hanin-Itzak, L. (2019). Sense of empowerment among school ICT coordinators: Personal, subject-area and leadership empowerment. *Tech Know Learn*, *24*(3), 401-417. https://doi.org/10.1007/s10758-017-9346-8

Avvisati, F. (2020). The Measure of socio-economic status in PISA: A Review and some suggested improvements. *Large-scale Assess Educ*, *8*(1), 1-37. https://doi.org/10.1186/s40536-020-00086-x

Bernacki, M. L., Greene, J. A., & Crompton, H. (2020). Mobile technology, learning, and achievement: Advances in understanding and measuring the role of mobile technology in education. *Contemporary educational psychology*, *60*, 101827. https://doi.org/10.1016/j.cedpsych.2019.101827

Bulut, O., & Cutumisu, M. (2018). When technology does not add up: ICT use negatively predicts Mathematics and Science achievement for Finnish and Turkish students in PISA 2012. *Journal of Educational Multimedia and Hypermedia*, *27*(1), 25-42. https://www.learntechlib.org/p/178514/

Burbat, R. (2016). Learner autonomy and ICT in foreign language teaching: Moving forwards or backwards? *PORTA LINGUARUM*(26), 37-51.

Bush, T. (2018). Transformational leadership: Exploring common conceptions. *Educational Management, Administration & Leadership*, *46*(6), 883-887. https://doi.org/10.1177/1741143218795731

Cai, S., Liu, E., Yang, Y., & Liang, J.-C. (2019). Tablet-based AR technology: Impacts on students' conceptions and approaches to learning mathematics according to their self-efficacy. *British Journal of Educational Technology*, *50*(1), 248-263. https://doi.org/10.1111/bjet.12718

Chiao, C., & Chiu, C.-H. (2018). The Mediating effect of ICT usage on the relationship between students' socioeconomic status and achievement. *The Asia-Pacific Education Researcher*, *27*(2), 109-121. https://doi.org/10.1007/s40299-018-0370-9

Demir, E. K. (2021). The Role of social capital for teacher professional learning and student achievement: A Systematic literature review. *Educational Research Review*, *33*, 100391. https://doi.org/10.1016/j.edurev.2021.100391

Gómez-Fernández, N., & Mediavilla, M. (2021). Exploring the relationship between Information and communication technologies (ICT) and academic performance: A Multilevel analysis for Spain. *Socio-economic Planning Sciences*, *77*, 101009. https://doi.org/10.1016/j.seps.2021.101009

González, A., & Paoloni, P.-V. (2015). Perceived autonomy-support, expectancy, value, metacognitive strategies and performance in chemistry: A Structural equation model in undergraduates. *Chemistry Education Research and Practice*, *16*(3), 64-653. https://doi.org/10.1039/c5rp00058k

Gubbels, J. C. G., Swart, N. M., & Groen, M. A. (2020). Everything in moderation: ICT and reading performance of Dutch 15-year-olds. *Large-scale Assessment in Education*, *8*(1), 1-17. https://doi.org/10.1186/s40536-020-0079-0

Hafen, C. A., Allen, J. P., Mikami, A. Y., Gregory, A., Hamre, B., & Pianta, R. C. (2012). The Pivotal role of adolescent autonomy in secondary school classrooms. *Journal of Youth and Adolescence*, *41*(3), 245-255. https://doi.org/10.1007/s10964-011-9739-2

Honarzad, R., & Rassaei, E. (2019). The Role of EFL learners' autonomy, motivation and self-efficacy in using technology-based out-of-class language learning activities. *The JALT CALL Journal*, *15*(3), 23-42. https://doi.org/10.29140/jaltcall.v15n3.170

Hu, X., Gong, Y., Lai, C., & Leung, F. K. S. (2018). The Relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A Multilevel analysis. *Computers & Education*, *125*, 1-13. https://doi.org/10.1016/j.compedu.2018.05.021

Li, S. C., & Choi, T. H. (2014). Does social capital matter? A Quantitative approach to examining technology infusion in schools. *Journal of Computer Assisted Learning*, *30*(1), 1-16. https://doi.org/10.1111/jcal.12010

Liu, M., Liu, L., & Liu, L. (2018). Group awareness increases student engagement in online collaborative writing. *The Internet and Higher Education*, *38*, 1-8. https://doi.org/https://doi.org/10.1016/j.iheduc.2018.04.001

Meng, L., Qiu, C., & Boyd-Wilson, B. (2019). Measurement invariance of the ICT engagement construct and its association with students' performance in China and Germany: Evidence from PISA 2015 data. *British Journal of Educational Technology*, *50*(6), 3233-3251. https://doi.org/10.1111/bjet.12729

Mora-Ruano, J. G., Heine, J.-H., & Gebhardt, M. (2019). Does teacher collaboration improve student achievement? Analysis of the German PISA 2012 sample [Original Research]. *Frontiers in Education*, *4*(85). https://doi.org/10.3389/feduc.2019.00085

Odell, B., Cutumisu, M., & Gierl, M. (2020). A Scoping review of the relationship between students' ICT and performance in mathematics and science in the PISA data. *Social Psychology of Education*, *23*(6), 1449-1481. https://doi.org/10.1007/s11218-020-09591-x

OECD. (2016a). *Does it matter how much time students spend on line outside of school? PISA in focus*. PISA, OECD. https://doi.org/10.1787/22260919

OECD. (2016b). *PISA 2015 results (Volume I)*. PISA, OECD. https://doi.org/10.1787/9789264266490-en

Park, S., & Weng, W. (2020). The Relationship between ICT-related factors and student academic achievement and the moderating effect of country economic indexes across 39 countries: Using multilevel structural equation modelling. *Educational Technology & Society*, *23*(3), 1-15.

Petko, D., Cantieni, A., & Prasse, D. (2017). Perceived quality of educational technology matters: A Secondary analysis of students' ICT use, ICT-related attitudes, and PISA 2012 test scores. *Journal of Educational Computing Research*, *54*(8), 1070-1091. https://doi.org/10.1177/0735633116649373

Sanders, M., & George, A. (2017). Viewing the changing world of educational technology from a different perspective: Present realities, past lessons, and future possibilities. *Education and Information Technologies*, *22*(6), 2915-2933. https://doi.org/10.1007/s10639-017-9604-3

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*(4), 765-781. https://doi.org/10.1037/a0012840

Szeto, E. (2020). School leadership in the reforms of the Hong Kong education system: Insights into school-based development in policy borrowing and indigenising. *School Leadership & Management*, *40*(4), 266-287. https://doi.org/10.1080/13632434.2019.1616172

Tam, V. C., Chan, J. W. W., Li, S. C., & Pow, J. (2018). Developing and managing school human capital for information and communication technology integration: A Case study of a school-based e-learning project in Hong Kong. *International Journal of Leadership in Education*, *21*(4), 447-461. https://doi.org/10.1080/13603124.2017.1318958

Wong, E. M. L., & Li, S. C. (2011). Framing ICT implementation in a context of educational change - A Structural equation modelling analysis. *Australasian Journal of Educational Technology*, *27*(2), 361-379.

Wu, H., Shen, J., Zhang, Y., & Zheng, Y. (2020). Examining the effect of principal leadership on student science achievement. *International Journal of Science Education*, *42*(6), 1017-1039. https://doi.org/10.1080/09500693.2020.1747664

Yalcin, S. (2018). Multilevel classification of PISA 2015 research participant countries' literacy and these classes' relationship with information and communication technologies. *International Journal of Progressive Education*, *14*(1), 165-176. https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1170065&site=ehost-live

Zhang, D., & Liu, L. (2016). How does ICT Use influence students' achievements in Math and Science over time? Evidence from PISA 2000 to 2012. *EURASIA Journal of Mathematics, Science & Technology Education*, *12*(9), 2431-2449. https://doi.org/10.12973/eurasia.2016.1297a

Romero-Hernandez, A., Gonzalez-Riojo, M., El Yamri, M., & Manero, B. (2022). The Effectiveness of a Video Game as an Educational Tool in Incrementing Interest in Dance among Younger Generations. *Educational Technology & Society*, *25* (4), 56-69.

The Effectiveness of a Video Game as an Educational Tool in Incrementing Interest in Dance among Younger Generations

Alejandro Romero-Hernandez*, Manuel Gonzalez-Riojo, Meriem El Yamri and Borja Manero

Complutense University of Madrid, Spain // alerom02@ucm.es // manuel.gonzalez@ucm.es // melyamri@ucm.es // bmanero@ucm.es

*Corresponding author

(Submitted October 4, 2021; Revised March 11, 2021; Accepted March 14, 2022)

ABSTRACT: The performing arts are currently in a critical situation worldwide. Various reports warn that the lack of audience. If we focus on dance, and especially folk dances, the situation is worse. In various countries and continents, folk dances are slowly disappearing. In Spain, we find evidence of the downward trend in terms of the number of attendees to performances of Spanish dance -an art form that is highly valued throughout the world. In a generation marked by technological advancements, the only way for classic performing arts to reach young audiences - or digital natives – is to speak the same language they use with new technologies. This paper presents a study in collaboration with the Spanish National Dance Company, carried out with 877 students (aged from 9 to 12) from 12 different schools in the community of Madrid, Spain. We designed a two-phase experiment. In the first phase, we separated the students into 3 groups: students who played a videogame called "Dancing a Treasure," those who received a workshop from professional dancers, and a control group. In the second phase that took place two weeks later, the participants attended to a real show of Spanish dance, and we studied how the previous educational approaches affected to the students increase of interest after the show. The experiment demonstrated that the videogame was, at least, as effective in increment interest about dance in younger generations as a workshop taught by expert dance professionals. Thus, in terms of scalability, the videogame is a better option because it can be applied with the same results to larger groups with no additional cost.

Keywords: Interest, Video games, Spanish dance, M-Learning, Serious games

1. Introduction

The performing arts are in a critical situation in Spain: smaller and smaller audiences attend shows, and their average age is increasing dangerously. If we focus on dance, according to the latest annual report (2017) published in Spain (SGAE, 2017), an art as long-lived as dance has had a decreasing number of spectators over the past ten years, with attendance reduced by half in this period (1.65 million in 2007 compared to 0.85 million in 2016). According to data collected in the latest survey of cultural habits in Spain, carried out every two years by the Spanish Ministry of Education (Ministerio de Educación, 2015), we can see that only 7% of those surveyed attended a dance show between 2014 and 2015. This situation is much worse if we focus on Spanish dance, so representative of Spanish popular culture. Only 17.9% of dance shows attended were specifically Spanish dance shows (1.25% attendees of the total number of people surveyed). These data represent a significant risk for this art form and all the companies and cultural assets related to it.

Moreover, this problem is not exclusive to Spain: the latest National Public Survey on Attendance, Participation & Engagement with the Arts in Ireland (Ireland Arts Council, 2018) shows that ballet is the art that people attend the least, with only 2% attendance in the last 12 months among those surveyed. At the same time, at folk dance performances, attendance barely reaches 9% of the public surveyed.

In countries on other continents, such as Chile (Ministerio de las culturas, 2017), they face similar problems to those already mentioned. Even though in said country arts such as ballet have slightly increased in the last year, statistics showed that numbers are still low. These are just a few recent examples of worldwide statistics that present a worrying situation for the survival of the dance arts in the mid- to long term.

One of the best ways to solve this problem seems to be by raising young people to be interest of this art. In many cases, in addition to ensuring the medium and long-term survival of this type of arts, it also directly influences public attendance due to the tow-along effect the youngest have on their families. Companies such as the Ballet Nacional de España (hereinafter BNE), the world's leading Spanish dance company (España, 2018), realize this, and they have taken action to become more accessible to children. The BNE has launched one main activity with that goal in mind: they have been organizing workshops in schools. These workshops are taught by the dance company staff and consist of several activities with the students to learn basic concepts of Spanish dance.

This activity is undoubtedly essential to reaching new audiences. However, there is a challenge: it's scalability. The workshop is an activity that can only be carried out in particular locations (schools) that are willing to include it in their schedule and, most importantly, it can only be taught in one place and at one time. This workshop is carried out by people from the organization itself, so they have to stop performing their functions in the company in order to free up time for the workshop, thus limiting their willingness to do so.

We live in the digital native era, which makes it difficult to reach audiences that receive information from many sources simultaneously (Piscitelli, 2008; Prensky, 2010). In order to reach these young people in the ecosystem in which they live, there is a method that tries to use so-called serious games (Michael et al., 2005). Besides, these types of videogames are used for raising awareness about other topics such as volcanic hazard education and communication (Mani et al., 2016) or cybersecurity (Hendrix et al., 2016), among others. Finally, there are already existing studies of useful techniques for the development of this type of tools (Wouters et al., 2017).

There is evidence of the possibility that new technologies can increase motivation towards different disciplines. There are studies that demonstrate the effectiveness of video games in motivating reading (Doran, 2010; Edwards, 2009). We can also find predecessor projects (Tongpaeng et al., 2018) where researchers applied serious games to increase the interest of young people in the arts, mostly in those with a higher probability of extinction due to their lack of spectators. Specifically, there are other videogames, based on classical plays, which have demonstrated the effectiveness of these tools in increasing young people's interest about art-related (Iglesias et al., 2013; Romero-Hernandez et al., 2018).

Video games can provide an ideal environment for increase interest for various reasons: (1) The youngest perceive video games as an element that is typical of their generation, so their acceptance of games as educational tools is very strong ; (2) They can offer a virtual environment that allows the player to become immersed in the narrative that is unfolding (Fernández-Vara, 2009a); (3) They allow the player to be the protagonist of the story (Blumberg et al., 2011), listen to the original music, see the animations recorded by professional dancers, etc.; and (4) Video games allow agency to the players, which it is not offered by a film or documentary (Fernández-Vara, 2009b).

To the researchers' best knowledge, most studies concerning the effectiveness of serious games are conducted within educational environments. Thus, they are limited to a pre/post experimental design that they compare with a control group. In the literature consulted, we have not encountered any experiments outside educational environments; their effectiveness was always measured directly in schools but not with subsequent experience in a real environment such as a show.

In addition, one of the main problems that we have detected in the studies carried out on the effectiveness of video games as motivational tools is that they only focus on immediate effectiveness, that is, the increase in interest that occurs immediately after playing the video games. However, we do not have knowledge of studies that analyze whether this increase in interest is sustained.

In this paper, we intend to measure the effectiveness of an educational video game as an increasing interest tool for increasing interest in Spanish dance. We pretend to measure any increase in interest at two moments in time: (1) Immediate effectiveness, increase in interest in the dance that is produced by the videogame immediately upon playing, and (2) Real effectiveness, or increased interest in dance produced by the video game and after having seen an actual dance performance.

- The first intervention takes place at the schools, considering the following educational points for comparison: (1) Video games, (2) Recurring workshops carried out by the BNE at schools and (3) Control group that does not participate in either activity.
- The second intervention takes place two weeks later and intends to show whether there is a difference in student interest in Spanish dance after having attended a live dance performance in each of the three groups described in the previous section.

In order to achieve this, we used the video game "Dancing a Treasure" (Romero-Hernandez et al., 2019), based on the previously mentioned book "Bailando un Tesoro" (Huidobro et al., 2016), developed in collaboration with the BNE.

The BNE has always said that “You can only love what you know.” Therefore, to get young people to love Spanish dance, they first needed to get to know it. The question we asked ourselves was: Can a videogame played before watching a real show make it more interesting?

To get a measure of effectiveness, the first step is to register the effect the tool has in comparison with other possible activities in the school context.

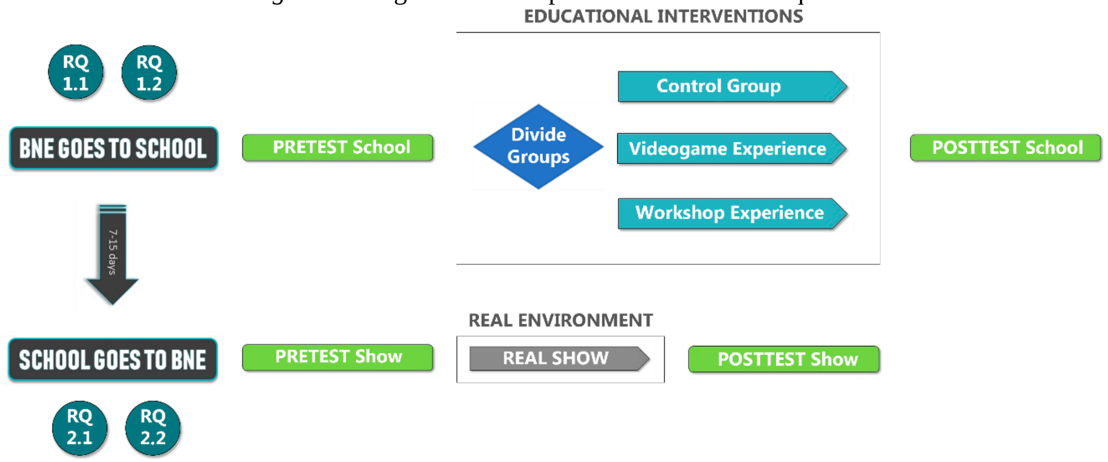
- **RQ1.1**. Is the videogame “Dancing a Treasure” effective in increment interest about Spanish dance?
  On the other hand, we are interested in learning the differences between this tool and those previously (mentioned above) used on other occasions by the BNE in the school context.
- **RQ1.2.** How effective is the video game “Dancing a Treasure” in producing immediate increased interest in Spanish dance in comparison with the traditional method used in the BNE’s workshops?

Following the process outlined before, the objective is to measure how effective the videogame is in increment interest in a real environment in the medium term. We wanted to measure how these same young people (after having previously attended the workshop, played the videogame or without any previous activity) perceived Spanish dance once they attended the show. On this basis, the following questions arose:

As we will demonstrate in the results section (see Section 4.2.1), attending a live Spanish dance show increases interest in Spanish dance among the audience.

- **RQ2.1.** Is that increment higher when the spectators have played the video game “Dancing a Treasure” before attending the show than when they do not do anything previously?
- **RQ2.2.** Is that increment higher in those who played the video game than in those who attended the traditional educational approach (workshop)?

Figure 1. Diagram of the experiment and research questions



This article is structured as follows. In section 2, we explain the methodology followed. In section 3, we present the results obtained in our research. In section 4, we discuss the results of the previous section. Finally, in section 5, we explain the conclusions, limitations, and future work.

2. Methodology

2.1. Experimental design

Quasi-experimental design is an empirical intervention study used to estimate the causal impact of an intervention on the target population (Gribbons et al., 1996; Miller et al., 2020). Following this, and as mentioned in section 1, the experiment consisted of two phases (see Figure 1): (1) the BNE goes to a school; we carried out this part of the experiment in the schools themselves, and (2) the school goes to the BNE; that is, when the students went to the BNE’s headquarters.

This experimental design supposes that there is a period of X days between the two phases. Because of scheduling conflicts on both the BNE's part and the schools', this period was not always the same, but we kept it within particular margins: 7 to 15 days at most between the first phase and the second one.

2.1.1. BNE goes to school (RQ1.1 and RQ1.2)

This subexperiment (that took place in the schools' classrooms) aims to test the immediate effectiveness of our game: How is the interest increasing in spanish dance just after playing the videogame? And, how is the interest increasing in comparison with the traditional educational approach (the workshop)? Thus, it will address RQ1.1 and RQ1.2.

2.1.1.1. Research design and data collection

In this first phase, students were divided into one of the following groups randomly: (1) the Control Group (CG), which received no instruction, (2) the Workshop Group (WG), which participated a workshop by the BNE staff, and (3) the Game Group (GG), which played the videogame Dancing a Treasure. For our study, the groups that resulted are statistically homogeneous and comparable (see section 4).

Due to the nature of the environment, we designed the activity to be completed in a single classroom session (50 mins approximately). We divided the activity as follows:

- **Pre-test (from now on PRETEST_School):** The researchers go into the three groups' classrooms and, without giving any information about the intervention so as not to raise students' enthusiasm in any case, they hand out some tests. The tests are anonymous (all students have received a code to identify their answers throughout the experiment). In these questionnaires, they are asked mainly about their interest in Spanish dance in general. We also collected demographic data such as sex, age, or the course they were in. This first test takes about 5 minutes. We will explain this tool in detail in section 3.3.2.

- **Intervention:**
    - The people in charge of organizing the **workshop** prepare the necessary materials to develop the WG activity, and the workshop starts, lasting about 40 minutes. This workshop is a purely practical experience for the students, where BNE members bring their work tools (shoes, castanets, etc.) and teach the essential company and Spanish dance concepts. Finally, the students can try out these tools, touch them, and watch firsthand, among other things, a live professional display of the rhythmic heel-stomping style of dance.
    - The researchers in charge of the **Game Group** distribute tablets (mostly Lenovo model TB-X103F) and headsets to each student. These tablets contain the previously installed game and have a shortcut on the desktop for easy access. As soon as all the students have the material ready, the game starts, lasting about 40 minutes.
    - The researchers in charge of the **Control Group** distribute the pre-test to the students and, later, they leave the classroom and allow the teachers to continue with the rest of the school day.

- **Post-test (from now on POSTTEST_School):** After finishing the activities, GG students and WG students complete a second test with the same questions as in the pre-test.

2.1.1.2. Data analysis

To answer question **RQ1.1** we performed a statistical analysis comparing the responses of the game group only. Through a *t*-student test we compared the increase of the responses mean between the PRETEST_School and the POSTTEST_School. However, to address question **RQ1.2** we will consider both game group and workshop group data. Through a one-factor ANCOVA test, we compare the difference in interest increments between the two groups above mentioned. Thus, we will use PRETEST_School as covariate and POSTTEST_School as dependent variable.

2.1.2. School goes to the BNE (RQ2.1 and RQ2.2)

After attending a real dance show, is the students' interest in spanish dance higher if they previously played a videogame (Bailando un tesoro)? Which interest increment is higher, those who played the videogame or those who attended the dance workshop? Thus, this subexperiment aims to answer RQ2.1 and RQ2.2. and took place at BNE's headquarters.

2.1.2.1. Research design and data collection

The second activity follows a similar scheme to the first one but with small changes. All these activities have the following scheme:

- **Pre-test (from now on PRETEST_Show):** We continue with the anonymous tests; students answer the interest questions regarding Spanish dance again. This test has been included so that the reader can check the evolution of the increase in interest produced by both educational interventions. As can be seen in the graph of the control group, the PRETEST_Show is influenced in a certain way by a variant of the Hawthorne effect. The simple fact of knowing that you are going to attend a real dance performance increases the interest of the subjects in it. For this reason, and because it is not within the scope of this paper, the data obtained in the PRETEST_Show are not used in this study.
- **Real Environment:** All three groups attend a private show at the BNE headquarters. This allows the students to experience Spanish dance in a real environment (see Figure 2).

Figure 2. Students at a BNE show

- **Post-test (from now on POSTTEST_Show):** After the show, the students answer the interest questions regarding Spanish dance again.

2.1.2.2. Data analysis

To address **RQ2.1** we will use game group y control group data. Again, we used the T-student test to compare the increment of interest (POSTTEST_Show - PRETEST_School). To answer **RQ2.2** we use the three groups' data (game group, control group, and workshop group). We used unifactorial ANCOVA to compare the interest increment among groups. PRETEST_School as covariate and POSTTEST_Show as the dependent variable.

2.2. Gaming preferences of students

This study (see Table 1) involved N = 877 students in 5th grade (9-12 years old) from 12 schools in the Community of Madrid (Spain). The gender ratio in the study group is 50.4% female to 49.6% male. The group's ages are between 9 and 12 years, with age 10 being the most frequent (72.1%). There were also 11-year-old students (26.2%), 12-year-old students (1.1%) and 9-year-old students (0.6%).

Regarding the schools, 47% of the students went to public schools, while 44% attended mixed/charter schools (funded both by public and private monies), and the remaining 9% were at strictly private schools. Both in terms of gender and school type, the sample is very similar to the population of students within the Community of Madrid in these courses (Madrid, 2019; Ministerio de Educación y formación profesional, 2019), so it can be considered representative of this area.

Table 1. Demographics of participants

Group	Number of participants		Age	Total
	Female	Male		
Game based instruction (GG)	96 (48.2%)	103 (51.8%)	10.22 +- 0.425	199 (22.7%)
Workshop based instruction (WG)	159 (52.0%)	147 (48.0%)	10.31 +-0.503	306 (34.9%)
Control Group (CG)	187 (50.3%)	185 (49.7%)	10.29 +-0.5	372 (42.4%)
Total	442 (50.4%)	435 (49.6%)	10.28 +-0.486	877

3. Materials and instruments

3.1. Dancing a Treasure: The Videogame

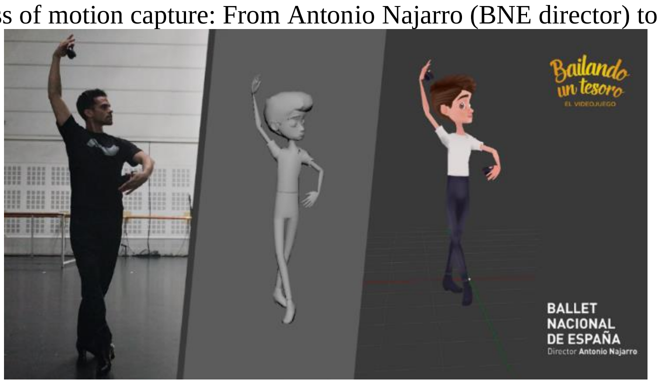
The videogame "Dancing a Treasure," described in detail in the article published in ICCE 2019 (Romero-Hernandez et al., 2019), is a videogame based on Spanish dance that aims to increment young people's interest about dance. The project was carried out with the collaboration of the BNE, and it is the first videogame based on Spanish dance that we are familiar with.

We based the videogame's narrative on the book of the same name "Bailando un Tesoro" (the name of the videogame is the literal translation from Spanish to English). It tries to give the best representation of the Spanish dance world (see Figure 3). Since a former dancer wrote this book, we can find parallels between this narrative and the beginnings of a dancer's career, making this story a perfect starting point.

The videogame does not focus on a particular performance. However, it includes various representations of the four styles that the BNE features (flamenco, folk, stylized dance and the bolero school). We designed it as such to disassociate the videogame from a specific show performed at a particular moment. The game could be applied to any of the functions that this company could be performing at any time.

The videogame has a very practical part based on Spanish dance. Dance is the star of the videogame. Throughout all levels, the main character stands in the middle of the screen. This character dances in all the styles of Spanish dance according to the narrative mentioned above. At the same time, the videogame has a purely rhythmic base. We designed a system of interactions for the player to mark the tempo and rhythm of the songs chosen for the videogame (from the BNE's repertoire). These interactions appear on both sides of the character and at the bottom of the screen so that the user can see the dance clearly while playing.

Figure 3. The process of motion capture: From Antonio Najarro (BNE director) to Dancing a Treasure



3.2. Questionnaires

For the development of this experiment, researchers created four questionnaires: (1) an experience pre-test in schools (***PRETEST_School***); (2) a post-test for the experience in schools (***POSTTEST_School***); (3) a pre-test to be filled in before the students' experience at the BNE headquarters (***PRETEST_Show***); and (4) a post-test at the end of the BNE experience (***POSTTEST_Show***). The main goal of those questionnaires is to measure the interest about Spanish dance. We could not find any validated tests, so we created ad-hoc tests for this experiment. We based these tests on other very similar ones from previous projects that we had as a reference. This is one of this experiment's clear limitations, and we will explain in more detail in the Limitations Section.

In the first pre-test, we collected demographic data, such as the participants' age and sex. This questionnaire asks about their knowledge of the BNE, Spanish dance and the number of times they attended a Spanish dance performance in the last year. The second part of the test consists of 4 questions (see Table 2) related to their interest of Spanish dance.

The tests, which the students completed during the different stages of the experiment, contain the same four questions about interest of Spanish dance. We based these questions on a Likert 7 scale (Joshi et al., 2015), and, in the end, we added them up to obtain the total result. Therefore, these sums can have a minimum value of 4 and a maximum of 28 in total.

Table 2. Questions and type of answers

Questions
Evaluate from 1 to 7: How much do you like to dance?
Evaluate from 1 to 7: How much do you like to dance Spanish dance?
Evaluate from 1 to 7: How much would you like to attend a Spanish dance performance?
Evaluate from 1 to 7: How much would you like to know more about Spanish dance?

After recolenting all these data, we carried out a data analysis that we explain in detail in section 4 of this paper. Figure 4 shows the four stages of the data analysis and the performed tests. To summarize, we performed an ANCOVA test to measure the educational approaches efficiency: (1) within the school environment (**PRETEST_School and POSTTEST_School**), and (2) in the overall experience (**PRETEST_School and POSTTEST_Show**).

Figure 4. Data analysis diagram

4. Results

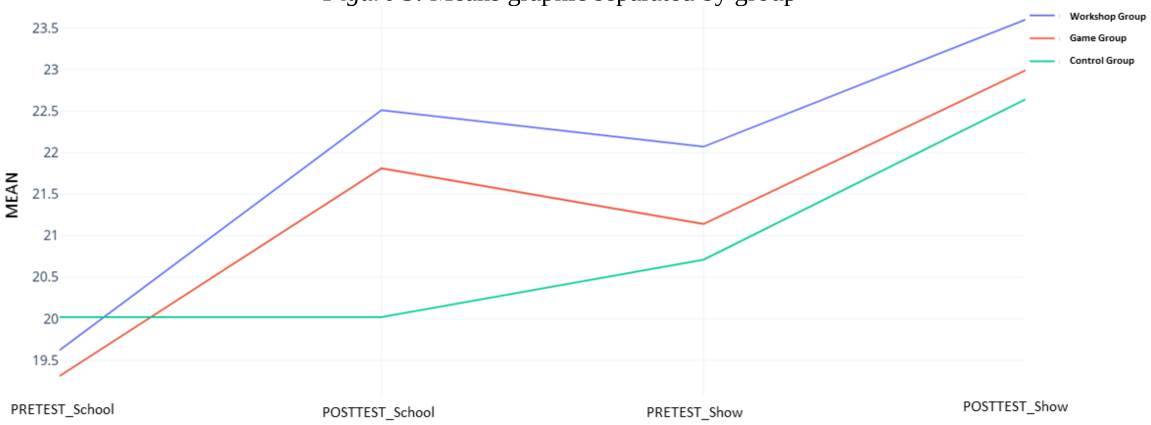
In the following section we will review the results of the experiments that were conducted. In order to understand the results better, we include below the general descriptive statistics table, which will serve as a guide for the entire section (see Table 3).

Table 3. Descriptive Workshop (WG), Control (CG) & Game (GG) Groups

Descriptive	Interest (I) in PRETEST_School		Interest (I) in POSTTEST_School		Interest (I) in PRETEST_Show		Interest (I) in POSTTEST_Show	
	Mean (4-28)	*SD*	Mean (4-28)	*SD*	Mean (4-28)	*SD*	Mean (4-28)	*SD*
Workshop Group	19.73	5.80	22.51	4.71	22.07	5.35	23.69	4.71
Game Group	19.34	5.55	21.81	5.05	21.14	5.08	23.13	5.05
Control Group	19.95	5.56	--	--	20.71	5.85	22.95	5.54

We also include a graphic which shows the evolution of the interest in the tree groups though the different tests (see Figure 5). The blue line represents the workshop group, the red one represents the game group and the green one represents the control group. Y-axis shows the interest average, while the x-axis shows the four data collection points. This allows the reader to see an overall summarize of the process.

Figure 5. Means graphic separated by group



4.1. BNE goes to School (RQ1.1 and RQ1.2)

In this section the objective is twofold: (1) To test if playing our videogame increases the interest in spanish dance (**Intra-group comparison**) (RQ1.1). To that aim, we compare the interest before and after playing the videogame; and (2) To compare the efficiency of both educational interventions (**Inter-group comparison**: Game vs. Workshop) (RQ1.2). Thus, we will compare the increment produced by the game with the one produced by the workshop.

4.1.1. Intra-group comparison (Game group and Workshop group)

To answer RQ1.1 we compared **PRETEST_School** (before playing the videogame) and **POSTTEST_School** (after playing the videogame) through a unifactorial *t*-test with the GG. The test gives us a statistically significant difference between the average of pre-test and post-test responses (*sig.* < .0001). More precisely, we observed a difference of 2.47 points between the average of the different tests (That is the increment of the means of interests. From now on, we will use the symbol ΔI to mention it) with a standard deviation of 3.61 (see Table 4).

We also checked the workshop effectiveness applying the same process. WG students displayed a statistically significant increase (see Table 4) in their interest of Spanish dance (*sig.* < .0001), with a difference of 2.78 points between the means of the two tests with a standard deviation of 5.39.

Table 4. t-Student test results intra-group comparison: Game Group and Workshop Group (POSTTEST_School /PRETEST_School)

GROUP	Mean	Std. Dev	T value	Sig.
WG	2.78	5.39	9.26	< .0001*
GG	2.47	3.61	9.75	< .0001*

Note. *p < .05.

4.1.2. Inter-group comparison (Game group vs. Workshop group)

To answer RQ1.2, we proceeded to compare the interest increment between the GG and the WG. To make this comparison, we had to confirm one premise: the two groups had to be homogeneous at the beginning (in PRETEST_School). For this purpose, we performed the equality of variances test (F = 1.10, *sig.* > .05). The test shows (see Table 5) no statistically significant differences between the groups, so their evolutions are comparable.

Table 5. Equality of variances in Game Group and Workshop Group (PRETEST_School)

	Method	*F* value	Sig
PRETEST_School	Folder F	1.10	.460

Note. *p < .05.

We perform the bifactorial ANCOVA to compare the average increase in students' interest using the POSTTEST_School and PRETEST_School variables and the group (GG or WG). This test showed no statistically significant differences (see Table 6) between WG and GG (*sig.* > .05). The interest increments are: WG (ΔI = 2.78, *SD* = 5.39) and GG (ΔI = 2.47, *SD* = 3.61).

Table 6. School ANCOVA test results inter-groups: Workshop group vs. Game Group (PRETEST_School/ POSTTEST_School))

	Mean Square	*F* value	*Sig.*
Workshop group vs. Game Group	27.66	1.10	.221

Note. $^{*}p < .05$.

4.2. School goes to BNE (RQ2.1 and RQ2.2)

Firstly, we need to check whether attending a dance show increases interest in dance. To this aim we used our Control group (CG), that had not performed any activity prior the experiment, and compared their interest before and after the show (see section 4.2.1. Control group analysis).

Once we demonstrated that attending to a dance show increases the interest towards this art, we need to know if playing a videogame before attending the same show could make that increment higher (RQ2.1). Finally, we want to measure the effectiveness of the videogame in comparison to the traditional educational approach (workshop)(RQ2.2). In the increment produced by attending a real show higher when the audience played the videogame in advance or when they attended the workshop? Section 4.2.2, Inter-group comparison (Game group vs. Workshop group vs. Control group) includes the results for both research questions.

4.2.1. Control group analysis

To determine whether assisting a dance show increases interest in dance, we must check if there is any significant difference between PRETEST_School and POSTTEST_Show of the CG. To analyse this, we apply a T-test for related samples. This test shows us that there is a significant statistical difference between PRETEST_School and POSTTEST_Show (see Table 7). There is a total increment ΔI = 3.00 (*SD* = 4.406) in the control group (CG).

Table 7. t-test results in control group (PRETEST_School/POSTTEST_Show)

	Mean	*SD*	*t* value	*Sig.*
Control Group (CG)	2.61	4.40	10.44	.000*

Note. $^{*}p < .05$.

4.2.2. Inter-group comparison (Game group vs. Workshop group vs. Control group)

Once we have the previous result, and to answer RQ2.1 and RQ2.2 questions, we must verify that the three groups start from the same point, i.e., PRETEST_School are homogeneous. To guarantee this, we performed a unifactorial anova (see Table 8) test (*F* = 0.72, *sig.* > .05). This test showed no statistically significant differences between the groups, so their evolution is comparable.

Table 8. Unifactorial ANCOVA test results in Workshop (WG), Control (CG) & Game (GG) Groups (PRETEST_School)

	Mean square	*F* value	*Sig.*
PRETEST_School	22.90	0.72	.480

Based on this homogeneity, we were able to perform the ANOVA Unifactorial test to see if there were statistically significant differences among groups (GG, WG or CG), by comparing POSTTEST_Show with PRETEST_School. This test showed statistically significant differences among the different groups (*sig.* < .05).

By separately analyzing the groups with the ANCOVA test (see Table 9), we observe statistically significant differences between WG and CG (*sig.* < .001), and between GG and CG (*sig.* < .05), but not between WG and GG (*sig.* > .05). The interest increments are: WG (ΔI = 3.96, *SD* = 4.222), GG (ΔI = 3.79, *SD* = 5.246), and CG (ΔI = 3.00, *SD* = 4.406).

Table 9. Show ANCOVA test results inter-groups: Workshop group vs. Game Group vs. Control Group (PRETEST_School/ POSTTEST_Show)

ANCOVA	WG	GG	CG
WG	X	.304	.033*
GG	.304	X	.000*
CG	.033*	.000*	X

Note. $^{*}p < .05$.

5. Discussion

As previously mentioned, questions RQ1.1 and RQ1.2 are centered on the comparison of the efficacy of the video game in the classroom (differences between PRETEST_School and POSTTEST_School). However, questions RQ2.1 and RQ2.2 focus on the analysis of the tool's influence when students attend a live performance of the BNE (differences between PRETEST_School and POSTTEST_Show).

5.1. RQ1.1. Is the videogame "Dancing a Treasure" effective in increment interest about Spanish dance?

The results obtained after carrying out the Unifactorial *t*-test (see Section 4.1) show a statistically significant increase in students' interest in Spanish dance after playing the video game. This means that the answer is yes: the video game increases students' interest in Spanish Dance.

These results were predictable, due to the increasing interest characteristics of the video game as demonstrated by similar studies applied to performing arts such as theater (Manero et al., 2015a; Romero-Hernandez et al., 2018). Thus, this result confirms that video games can also be used to increase interest in other performing arts.

However, when analyzing these results, one should not ignore the effect that the novelty of playing a video game during class time can have on students. This effect, which can also be produced during other educational interventions such as the workshop, would require a separate study to demonstrate which part of the increase in interest is due to the educational intervention and which part is due to the novelty.

5.2. RQ1.2. How effective is the video game "Dancing a Treasure" in producing immediate increased interest in Spanish dance in comparison with the traditional method used in the BNE's workshops?

The results obtained after carrying out the bifactorial ANCOVA test to compare interest between the GG and WG groups do not show statistically significant evidence for concluding that one tool is more effective than another.

This finding differs from those obtained previously in other studies (Manero et al., 2015). In Manero's paper, researchers found statistically significant differences between playing a video game and attending a workshop; specifically, an actor-led class was more efficient than a video game. In this study we expected a similar result which is to say that the workshop was more efficient than the video game.

On the other hand, researchers' observations led us to believe that our videogame could have worked better than the one in Manero's experiment and outperform the workshop. Within the GG, we observed a pattern in how they interacted with the game: they started with silent calm and moved to noisy excitement. This excitement led to creating a competitive environment to see who would go the farthest in the game. At the end of the experience, we observed more excitement in the GG students than in the WG students, which was ultimately not reflected in the questionnaire data.

At the end, it seems that our videogame worked better than those used in previous studies. It is possibly due to various circumstances.

- The first is that we created a narrative that allowed a better immersion in the dance world than the one used in the aforementioned study. The story and characters design resulted adequate. According to the researchers' observations, the children started imitating the main characters movements. This leads us to believe that they developed a strong bond with the game's characters.
- Following the same idea, the design of the video game levels may have played a role in their impact on efficiency. Within the GG, we observed a pattern in the way they interacted with the game: they started out

quietly calm and swapped to a noisy excitement. This excitement led to creating a competitive environment to get the highest level in the game. Definitely, after playing the videogame, the classroom atmosphere was fun.

- Another possibility is that this type of technology is better adapted to a very visual art such as dance. Music and dance movement are very adaptable to audiovisual media.
- On the other hand, the platform used for playing video games is different than the PCs used by Manero et al. (2015). In this study mobile devices were used. It is possible that these devices increase students' interest more greatly.
- Finally, this video game doesn't use any particular dance performance. While Manero et al. (2015) created a video game based on a concrete theatrical play, "Bailando un tesoro" is not based on a specific work but on the life of a dancer.

5.3. RQ2.1. Is that increment higher when the spectators have played the video game "Dancing a Treasure" before attending the show than when they do not do anything previously?

Although not the objective of our research, it is very interesting to note that according to our results, the attendance of a live performance of the National Ballet of Spain causes students' interest in Spanish dance to increase significantly (see Table 7). This result was to be expected, but we have not been able to find any study that proves it. In fact, as we see in Figure 5, the mere fact of leaving school and going to the BNE headquarters increases interest (as can be seen in the Control Group students). This is probably due to the excitement generated by breaking the usual class routine to go somewhere else, especially to see a show.

That said, regarding question RQ2.1, the answer is yes. The ANCOVA test demonstrates that those students who played the "Dancing a Treasure" video game before attending the show (Game Group) showed more of an increase in interest in Spanish dance than those who only attended the show (Control Group).

This may be because familiarity with the different styles and dances in advance makes what is being seen in the real show more comprehensible and, therefore, the message is much clearer. As we stated in the introduction: "You can only love what you know." These results are the best expression of it.

This result shows the possibility of using video games for purposes other than pedagogical projects. The familiarity effect produced by video games can make a show more enjoyable: when the viewer is familiar with the plot in a classical play, or when a spectator reads a libretto before an opera performance, the experience is improved.

5.4. RQ 2.2. Is that increment higher in those who played the video game than in those who attended the traditional educational approach (workshop)?

There is no statistically significant difference between Game Group and Workshop Group. It means that the "show effect" is the same either the students played the game or attended the workshop in advance. Both groups increase the interest in Spanish dance in a similar way after the show. As seen before, attending to a real show increases the interest in Spanish dance, but our results showed that the increasing is higher when the participants previously played our videogame or attended the workshop. Both approaches are equally efficient.

This does not coincide (as in RQ 1.2) with previous studies. In fact, we consider that the reasons why the game works better than we expected coincide with what is written in RQ1.2. However, in this case we feared the deception effect that those who played the video game could experience. Something like: "This show is not what they sold me in the video game." This effect did not occur, and the increase in interest after the show is statistically the same in the Game Group and in the Workshop.

Our starting hypothesis was that the workshop would be more efficient than the video game. We thought that the fact that real dancers were working with them during the workshop would make real show attendance more efficient as a increaser of the interest. As shown in Figure 5, the results of the GG and the WG are parallel throughout the experience: both groups start with a similar level of interest, experience a strong increase in interest after the educational intervention, experience a slight decrease in interest over the following two weeks, and a new increase after the show. And since there are no statistically significant differences, we could say that both interventions are interchangeable in terms of efficiency.

In addition, we asked the participants to evaluate the experience (EE) from 1 to 7 at the end of the experiment (after attending the real show). Answers showed that those in the workshop group evaluated the experience (EE = 6.61, *SD* = 0.885) higher than those in the game group (EE = 6.39, *SD* =1.338). The participants of the control group got the lower result (EE = 6.17, *SD* = 1.313). These results, not included in the results section to improve clarity, led us to think that the human factor of the workshop is key in the perception of the experience. It seems that having attended the workshop or played the videogame improves the perception of a real dance show. In other words, the workshop or the videogame make the real show more fun.

We would also like to point out that the decline in interest that occurs during the two weeks after the educational intervention is similar in both groups. Although it is not within the scope of this paper, we consider that it would be interesting to investigate the decrease in interest that occurs over time in function of the educational intervention used.

6. Conclusions and limitations

This study presents the results of using the videogame "Dancing a Treasure" as a tool to increment students' interest of Spanish dance. After collecting and analyzing the data obtained in the experiment, we conclude that the videogame increases the interest of Spanish dance and improves students' perception when they attend a dance performance. The videogame ended up being as effective as a workshop given by the BNE staff. This project corroborates previous studies that claimed that a videogame is an effective tool to increment interest towards the arts. It also reveals a new use for this type of videogames: they can enhance the experience of attending a real performance.

Demonstrating that a videogame could be as effective as a workshop is good news since the latter has a higher long-term cost: the dancers' travel expenses to get to the schools, their salary, etc. At the same time, these workshops have some significant limitations when it comes to reaching many people (its main goal) since BNE members have limited availability and can only be in one location at a time. The videogame's benefits regarding these restrictions are numerous. The videogame only has an initial outlay (with significant support from public organizations due to the widespread acceptance of new technologies in the classrooms). It can also be played on as many devices as desired simultaneously, thus eliminating many of the disadvantages of other educational approaches. Our recommendation, based also in the researchers' observations, would be to include videogames in educational programs (along with the traditional approaches) to strengthen them.

Unlike other previous studies, in which researchers based the tools on a specific theater play (Iglesias et al., 2013), the videogame "Dancing a Treasure" is a generic tool of Spanish dance. It is not tied to a specific choreography. This means that the videogame may be useful before any Spanish dance show, either by the BNE or by other companies, thus avoiding the cost of developing a new videogame for each show. This dramatically reduces the costs that these types of tools can involve. School campaigns usually carry out cultural outings every year. Since the videogame is free to download, the cost of implementing a videogame that students can play before going on the field trip would be zero (if the necessary equipment is available) and would enhance the experience of going to a show.

There are some limitations to this study. First, we conducted the tests in 12 schools that accepted the invitation to participate, from the many schools that were invited. This may lead to some bias in the data, as the participating schools have often shown interest in cultural activities on other occasions, through regularly programmed events throughout the school year. Furthermore, we only conducted the study in the Community of Madrid, so the data come from a specific region. To tackle this limitation, we will include schools with less interest in these activities and schools from other locations, ideally from the whole of the Spanish territory. The second limitation is the type of tests used. We didn't find any validated tool in the previous work for the proposal of this project, so we had to develop an ad-hoc tool based on similar projects (Hernandez et al., 2016; Manero et al., 2015a). These have a small number of items (4-7 items) because of the reduced time available for the experiments' development. Henceforth, it would be interesting to develop and validate a tool with the support of psychologists and statisticians focused on the audience and the objective sought. At the same time, the videogame has a series of limitations due to the limited budget available for its development. With a bigger budget, a higher number of professional graphic designers, and the recording of dances with more advanced technology, we could have a much more effective videogame.

The present study raises the possibility of numerous future lines of research. First, we aim to analyze the collected demographic variables to see if there are any findings of interest. Second, according to researchers' observations, at the end of the experience, students tried to imitate the main character's movements in the videogame. Thus, we aim to to develop a virtual reality experience that allows students to put themselves in the shoes of a dancer. And third, to check whether everything learned in this project can be extrapolated to the development of other arts or skills. We maintain the hypothesis that exercises related to dance or theatre can help to improve bodily expression, and we can apply this to the improvement of soft skills such as public speaking.

Finally, as already mentioned, the BNE has been organizing family performances in recent years, with discounts for families with children. At the start of this project (2016), BNE management itself told us that, to ensure that the theatre had an acceptable attendance rate, they invited people from the organization or held drawings through social networks. At the end of this project (2019), the same sources assured us that the tickets for these special performances were sold out weeks before performance day. We do not have empirical data to support that this project influenced the attendance rates, but we are sure that it helped in some way, which gives us hope to think that new technologies can be a solution to fill theaters again.

Acknowledgment

We would like to thank each and every one of the people who have collaborated in this project, both in the video game development and in the experiments' development. First of all, we would like to thank the company Ballet Nacional de España for their support in all the project phases, especially Belén Moreno for her original idea and for the effort in each project phase. On the other hand, we would like to thank the BNE director, Antonio Najarro, for his support and for lending himself to the video game dances recording. We would also like to thank Ann Kreis, BNE benefactor and of this project, for her great economic support and for her enthusiasm with the project. We would also like to thank Dr. Clara Fernández-Vara and Matthew Weise for their help with the narrative and design of the video game. Finally, we would like to thank Ricardo García, a statistician from the Complutense University of Madrid, for his great support in analyzing the experiment results.

This project has been partially funded by BBVA foundation (ComunicArte project: PR2005-174/01), and Ministry of Science, Innovation and Universities of Spain (Didascalias, RTI2018-096401-A-I00).

References

Blumberg, F. C., & Altschuler, E. (2011). From the playroom to the classroom: Children's views of video game play and academic learning. *Child Development Perspectives, 5*(2), 99–103.

Doran, J. S. (2010). Using video game playing to increase student motivation to read (Unpublished doctoral dissertation). Ohio Dominican University, Columbus, OH. http://rave.ohiolink.edu/etdc/view?acc_num=oduhonors1272911814

Edwards, B. (2009). Motivating middle school readers: The Graphic novel link. *School Library Media Activities Monthly, 25*(8), 56–58.

España, B. N. de. (2018). *Ballet Nacional de España* [National Ballet of Spain]. https://balletnacional.mcu.es/

Fernández-Vara, C. (2009a). *Play's the thing: A Framework to study videogames as performance*. MIT Libraries. http://hdl.handle.net/1721.1/100276

Fernández-Vara, C. (2009b). The Secret of monkey island: Playing between cultures. In *Well Played 1.0: Video Games, Value and Meaning* (pp. 331–352). ETC Press.

Gribbons, B., & Herman, J. (1996). True and quasi-experimental designs. *Practical Assessment, Research, and Evaluation, 5*(1), 14. https://doi.org/10.7275/fs4z-nb61

Hendrix, M., Al-Sherbaz, A., & Victoria, B. (2016). Game based cyber security training: Are Serious games suitable for cyber security training? *International Journal of Serious Games, 3*(1), 53–61.

Hernandez, A. R., Perez, C. D.-F., Riojo, M. G., Morata, A. C., & Manero, B. (2016). Computers or tablets? Choosing devices for educational videogames. *Proceedings of the Games and Virtual Worlds for Serious Applications (VS-Games), 2016 8th International Conference on* (pp. 1–8). IEEE. https://doi.org/10.1109/VS-GAMES.2016.7590342

Huidobro, A., & Palacios, M. (2016). *Bailando un Tesoro* [Dancing a treasure]. Institution Nacional De Artes Escenicas Y De Teatro. Ministerio de Cultura Secretaría General Técnica. Centro de Publicaciones.

Iglesias, B. M., Fernández-Vara, C., & Fernández-Manjón, B. (2013). E-learning takes the stage: From La Dama Boba to a serious game. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje, 8*(4), 197–204.

Ireland Arts Council. (2018). *Attendance, Participation & Engagement with the arts in Ireland 2018*. https://www.artscouncil.ie/uploadedFiles/wwwartscouncilie/Content/Arts_in_Ireland/Strategic_Development/Arts-Council-National-Survey-2018.pdf

Madrid, C. de. (2019). *Datos y Cifras de la educación* [Education facts and figures]. https://www.comunidad.madrid/sites/default/files/doc/educacion/sgea_datosycifras_presentacion_19_20.pdf

Manero, B., Torrente, J., Serrano, Á., Martínez-Ortiz, I., & Fernández-Manjón, B. (2015). Can educational video games increase high school students' interest in theatre? *Computers & Education, 87*, 182–191. https://doi.org/10.1016/j.compedu.2015.06.006

Mani, L., Cole, P. D., & Stewart, I. (2016). Using video games for volcanic hazard education and communication: an assessment of the method and preliminary results. *Natural Hazards and Earth System Sciences, 16*(7), 1673–1689.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade. https://dl.acm.org/doi/abs/10.5555/1051239

Miller, C. J., Smith, S. N., & Pugatch, M. (2020). Experimental and quasi-experimental designs in implementation research. *Psychiatry Research, 283*, 112452. https://doi.org/10.1016/j.psychres.2019.06.027

Ministerio de Educación, C. y D. (2015). *Anuario de estadísticas culturales* [Cultural statistics yearbook]. https://www.culturaydeporte.gob.es/dam/jcr:5df20471-770d-4261-a906-11eb15f4b3e4/anuario-de-estadisticas-culturales-2015.pdf

Ministerio de Educación y formación profesional. (2019). *Datos y cifras del Curso escolar 2018/2019* [Facts and figures of the 2018/2019 school year]. https://sede.educacion.gob.es/publiventa/descarga.action?f_codigo_agc=19183

Ministerio de las culturas, las artes y el patrimonio. (2017). *Estadisticas culturales* [Cultural statistics]. https://www.cultura.gob.cl/publicaciones/estadisticas-culturales-2017/

Piscitelli, A. (2008). Nativos digitales/Digital natives. *Contratexto*, (16), 43–56.

Prensky, P. M. (2010). *Nativos e inmigrantes digitales* [Digital and immigrants natives]. Cuadrenos SEK 2.0. https://www.marcprensky.com/writing/Prensky-NATIVOS%20E%20INMIGRANTES%20DIGITALES%20(SEK).pdf

Romero-Hernandez, A., Gonzalez-Riojo, M., El-Yamri, M., & Manero, B. (2019). Dancing a treasure: A Videogame to motivate young audiences toward Spanish dance. *Proceedings of the ICCE 2019 - 27th International Conference on Computers in Education, Proceedings (Vol. 1)*. https://apsce.net/icce/icce2019/proceedings/paper_93.pdf

Romero-Hernandez, Alejandro, Riojo, M. G., Díaz-Faes-Pérez, C., & Manero-Iglesias, B. (2018). The Courtesy of Spain: Theater for the new generations. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje, 13*(3), 102–110.

SGAE. (2017). *Anuario de la SGAE*. http://www.anuariossgae.com/anuario2017/home.html

Tongpaeng, Y., Mahamud, A., & Sureephong, P. (2018). Performing arts student attitude towards dance games as a Thai dance training tool. *Proceedings of the 2018 International Conference on Digital Arts, Media and Technology (ICDAMT)* (pp. 202–206). IEEE. https://doi.org/10.1109/ICDAMT.2018.8376524

Wouters, P., & Van Oostendorp, H. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In *Instructional techniques to facilitate learning and motivation of serious games* (pp. 1–16). Springer.

Knight, S., Matuk, C., & DesPortes, K. (2022). Guest Editorial: Learning at the Intersection of Data Literacy and Social Justice. *Educational Technology & Society*, *25* (4), 70-79.

Guest Editorial: Learning at the Intersection of Data Literacy and Social Justice

Simon Knight[1], Camillia Matuk[2*] and Kayla DesPortes[2]
[1]University of Technology Sydney, TD School, Centre for Research on Education in a Digital Society, UTS Broadway St, Broadway, NSW 2007, Australia // [2]New York University, Steinhardt School of Culture, Education and Human Development, Educational Communication and Technology, 370 Jay St., Brooklyn, NY 11201, United States // simon.knight@uts.edu.au // cmatuk@nyu.edu // kayla.desportes@nyu.edu
[*]Corresponding author

ABSTRACT: With growing awareness of, and attention to, the potential of data to inform decisions across contexts, has come an increasing recognition and need to develop data literacy strategies that support people to learn to be critical of data, given this consequential nature of data use (and abuse). To achieve a just society, inequities in both capacity for data literacy, and the applications of data in society, must be addressed. A key aim is to create learning experiences that engage learners with issues of power and inequity, including those typically marginalized by data literacy education. In this way, data literacy and social justice learning goals are intertwined, and mutually supportive, in developing data literacy in learning about, through, and for social justice. This special issue assembles five empirical studies on learning at the intersection of data literacy and social justice, and that illustrate various approaches to intertwining data science and social justice learning goals. They moreover highlight the importance of the learning sciences as a perspective for understanding how people learn in specific contexts of data justice. This essay reflects on themes raised by these contributions, and offers a framework for conceptualizing the intersections between the learning of data literacy and justice. In particular, we draw on existing distinctions between "reading" and "writing the world," and propose a mapping of data literacy justice activities from data comprehension to participation, and from thin to thick justice.

Keywords: Data justice, Numeracy, Mathematics education, Civics education, Equity

1. Why data literacy and justice?

Recent global events have shown the role that data can play in both causing and answering social injustices. The explosion of Big Data into all aspects of society, from its use in tracking physical and online activity, to its role in making personal and governmental decisions; as well as technological advances with respect to generating and accessing data have, on the one hand, empowered people to participate in society through the generation and use of data. Correspondingly, we have witnessed a growing thirst for data to inform both local and global strategies for addressing society's grand challenges, from climate change to pandemics (Lee & Campbell, 2020; Pennington et al., 2020). However, on the other hand, unawareness of how actions produce traces of data, and to whom these are available, have made communities more vulnerable to abuse by those who would commodify that data (Pangrazio & Selwyn, 2019; Raffaghelli, 2020). Similarly, lack of awareness and attention to the inherent biases of algorithms that predict and categorize people using big data have perpetuated racial, gender, and economic inequities (Buolamwini & Selwyn, 2018; Noble, 2018; O'Neil, 2016; Vakil & Higgs, 2019). Beyond question, the destructive impacts of unregulated uses of algorithms and big data have been felt across sectors including education, law enforcement, and healthcare (O'Neil, 2016).

The inaugural "Data Justice" conference, held in 2018 in recognition of the intersection of datafication and social justice concerns, noted "a need to position data in a way that engages more explicitly with questions of power, politics, inclusion and interests, as well as established notions of ethics, autonomy, trust, accountability, governance and citizenship" (Dencik et al., 2019, p. 874). Others have similarly advanced the notion of data justice, that is, "fairness in the way people are made visible, represented and treated as a result of their production of digital data" (Taylor, 2017, p. 1), an idea that is relevant to various realms of justice, including racial justice, economic justice, environmental justice, and spatial justice. More specifically, data justice is characterized by (1) visibility in terms of privacy and representation, (2) engagement in terms of autonomy in using data technologies, and accessing the benefits of data, and (3) non-discrimination, manifested by the challenge and pre-emption of bias and discrimination (Taylor, 2017).

Yet, learning to engage with data justice assumes a certain level of understanding the nature of data, how data are produced, and how data can be used (Raffaghelli, 2020). A degree of quantitative skills has long been recognized as central for successful participation in society (Gal, 2002; PIAAC Numeracy Expert Group 2009; Steen 2004; Engel, 2017), important in the workplace and valued by employers (Durrani & Tariq 2012; FitzSimons & Coben

ISSN 1436-4522 (online) and 1176-3647 (print).

2009). It is also noted that attention to the intersection of mathematics education and social justice is not new (e.g., Frankenstein, 1983). For example, researchers have explored ways to support the learning of mathematics through learners' critical examination of injustices toward low-income and minoritized communities, as evidenced in data that reveal disproportionate police stops and inequitable housing prices between communities (Gutstein, 2003; Gutstein, 2006).

However, in response to the injustices brought about—and brought to light—by the exponential growth in data, definitions of data literacy in education have expanded over the years. Once focused on quantitative and procedural skills, such as manipulating data sets, selecting and applying appropriate analyses, and making data-based inferences and arguments (See this overview of United States K-12 standards related to data literacy, https://tinyurl.com/dataliteracy-gdoc, Common Core State Standards Initiative, 2010; Franklin et al., 2007; NGSS Lead States, 2013), data literacy now encompasses abilities to critically reason with and about data; that is, to evaluate data and arguments with attention to the contexts in which data are produced and used, and to the people that are impacted (American Statistical Association, 2016).

This critical approach to data literacy is broadly relevant to, and urgently needed by all, a point reflected in several notable mainstream publications (e.g., D'Ignazio & Klein, 2020; Gigerenzer, 2014; Kahneman 2011; Levitin 2016; O'Neil, 2016). As citizens, we engage data literacy across (1) our general understanding of data in society regarding, for example, the economy ("data thinking"), (2) in our own ethical engagements with data ("data doing"), and (3) in more proactive data use, protection, and support for other's engagement with ("data participation") (Carmi et al., 2020; Yates et al., 2001). The creation, collection, and interpretation of data for specific purposes (and, where repurposed), is shaped by the social and material contexts in which the data is collected and used, reflecting features of power and culture (Stone, 2018; Van Wart et al., 2020). People must learn to be critical of data given the consequential nature of its use (and abuse), through questioning the socio-historical context of that data.

Yet, even the question of preparing a data literate citizenry is entangled in social justice issues. Notably, there remain inequitable outcomes in data-related disciplines (as across STEM) including low percentages of women and BIPOC (Black, Indigenous, and People of Color) individuals earning engineering, computer science, mathematics, and statistics degrees (NCSES, 2021), with some inequitable outcomes within the disciplines beginning at the school level (Quinn & Cooc, 2015). Thus, failing to address inequities in data literacy risks further perpetuating disparities in health, wealth, education, and employment. It follows that we cannot achieve a just society without addressing both the inequities in data literacy capacity, as well as the inequities caused by data's applications.

Building on these ideas, this special issue advances the significance of applying a learning research lens to understand the relationships between data literacy and data justice. This idea stands on the premise that, on the one hand, building data literacy can empower individuals and communities to produce, interpret and use data; and on the other, that building data literacy alone is not a panacea to social injustice (Philip et al., 2013). Rather, a focus on social justice is necessary for learners to engage critically with the underlying biases inherent in the conceptualization, operationalization, and interpretation of data projects (Kitchin, 2014); and to see data as models open to change, interpretation, and reinterpretation (Pangrazio & Selwyn 2018), rather than static and objective truths. Part of achieving this goal is through the design of learning experiences that support learners typically marginalized by data and data literacy education in gaining the statistical and mathematical skills to work with data; and that engage learners with issues of power and inequities in society with data. Thus, powerful learning experiences that support data literacy and social justice are ones in which statistical and social justice learning goals are intertwined and mutually supportive of one another.

This special issue assembles five empirical studies on learning at the intersection of data literacy and social justice, and that illustrate various approaches to intertwining data science and social justice learning goals. They moreover highlight the importance of the learning sciences as a perspective for understanding how people learn in specific contexts of data justice, including high school classrooms (Khan et al., 2022; Louie et al., 2022), out-of-school programs for economically disadvantaged students and students of color (Arastoopour Irgens et al., 2022; Khan et al., 2022); a training program for marginalized data workers (Shapiro et al., 2022), and a professional development workshop for students in higher education (Bhargava et al., 2022). Together, these contributions demonstrate approaches to data literacy via learning about, through, and for social justice.

2. Learning at the intersection of data literacy and data justice

By way of framing the contributions in this issue, this essay provides a conceptualization of the relationship between data literacy and social justice into what we refer to as *data literacy justice*. Existing conceptualizations of social justice oriented mathematics education identifies two interrelated goals to be pursued in tandem, and that Gutierrez (2009) has distinguished as “learning to play the game” vs. “learning to change the game.” By “playing the game” learners develop the foundational domain competencies (e.g., working with data tables and graphs, selecting appropriate analyses), necessary to eliminate inequities in academic performance based on demographics. Meanwhile, by “changing the game,” learners “read the world” (Gutstein, 2006), using these domain competencies to develop a critical understanding of social phenomena. Learners may further use these competencies and understandings to “write the world” (Gutstein, 2006), that is, to evoke social change through awareness, understanding, and advocacy (Xenofontos et al., 2021).

Rather than two opposing ends of a spectrum, these goals are deeply integrated. That is, data literacy pedagogy intended to simultaneously develop social justice literacy must go beyond only building foundational mathematics competencies. It must guide learners to recognize and respond to how data reflect the biases inherent in the systems that create and use them (Kitchin, 2014), how they are subject to multiple (mis)interpretations (Pangrazio & Selwyn 2018), and how they can echo power relations in society (Van Wart, Lanouette & Parikh, 2020). In essence, data literacy justice must put equal emphasis on preparing learners to both “play the game” and “change the game” (Gutierrez, 2009), by both “reading” and “writing the world” (Gutstein, 2003; Gutstein, 2006).

While the goals of data literacy justice education are relatively established, *how* to approach these goals through the design of learning experiences is less clear. The contributions in this special issue build on the existing literature in mathematics and civics education, participatory design, arts-based methodologies, and culturally sustaining pedagogy. The authors explore new ways to support data literacy justice by enabling learners to choose how they engage with data in personally meaningful ways, and by examining how creative representation can involve audiences in learners’ data work around social and political phenomena.

Below, we elaborate on three features common across the learning designs described in this issue: (1) Centering learners from historically marginalized populations in data work that extends from their lived experiences as individuals, and as members of their communities; (2) humanizing data through its creative representation, involving visual, embodied, and multimodal practices to reason about humanistic perspectives; and imagining new futures that integrate ethical and caring representations of their communities; (3) using socially situated data artifacts to “write the world.” We then propose a way of mapping data literacy justice activities from data comprehension to participation, and from thin to thick justice.

2.1. Centering learners in data work of personal and social relevance

One of the ways that the contributions in this special issue address the goals of data literacy justice is in their explicit focus on social issues caused by, or made visible by data; and, in their active inclusion of members of the communities most impacted by those issues.

In Louie et al. (2022) for example, high school students from economically disadvantaged communities whose first language was not English, engaged in structured data investigations of income equality and immigration. The work demonstrates how educators can develop students’ individual interests in data, their social and political consciousness, and their development of data literacy skills through combining real-world data from the U.S. Census Bureau with inquiry scaffolds directed at guiding sociopolitical investigations and discussions. For example, students investigated, “How much income inequality exists between males and females, and does education help to explain the gender wage gap?” and “Are immigrants less likely than U.S.-born individuals to participate in the labor force, before and after controlling for education?” Further, their work identifies how CODAP (Common Online Data Analysis Platform) supported the development of learners’ data literacy skills during these investigations through access to person-level data, engagement in multivariable analysis, comparing measures of center, and working across various canonical data representations.

Meanwhile, Shapiro et al. (2022) explore an approach to building data science capacity among members of populations historically excluded from STEM through employment. They describe a program that employs participants as data workers, and trains them in basic data science competencies through work on data relevant to their local communities. Through data projects submitted by for- and non-profit organizations, participants develop skills in data cleaning, formatting, and labeling, while also critically examining and representing issues that directly and personally impact them (in this case, data on policing within Black communities). Notably, the authors emphasize the importance of designing workplaces that actively promote the histories and perspectives of those systematically excluded and marginalized. They highlight the need to build data science knowledge upon learners' lived experiences, and encourage the use of intuitive multimodal, embodied approaches to sensemaking and communicating with data.

Similarly, in Kahn et al. (2022) high school students worked with local data from their community as they investigated the impact of the lottery in their communities. The work demonstrates the power of educational designs that scaffold learners to take a multimodal approach using mathematical, quantitative, and qualitative data reasoning. Coupling interviews with numerical data to understand social phenomena within a community promoted pluralistic perspectives rather than the monolithic views often used to frame minoritized communities. Learners found that individual reasons for playing the lottery ranged from addiction, to wanting to fund their children's education. Gathering various kinds of data from multiple sources allowed students to describe a holistic picture of the lottery's impact that accounted for the personal experiences of those implicated.

Arastoopour Irgens et al. (2022) go beyond learners' basic data science skills to support their reasoning about the uses of, and implications of emerging technologies in data science. Youth participants in an after-school program learned about algorithmic bias by making their own machine learning robots, and reflecting on how algorithmic bias in everyday technologies, such as facial recognition, can contribute to perpetuating larger scale racial and gender inequities. The work shows how youth can be centered in these explorations by giving them the capability to train their own algorithms, test them, and evaluate how and why they break down. By using these learner explorations as starting points, it provides them with first-hand experiences of how biased and faulty results can be programmed into these systems as they learn about data-driven computing systems in the real world.

These participatory approaches moreover build on learners' cultural values and interests to promote sustained engagement (DiSalvo & DesPortes, 2017; Druin, 2002). More specifically, they demonstrate the value of using learners' and their communities' own lived experiences as reflected in their focal data, as starting points for building the engagement necessary to develop comprehension, as well as the relatable contexts necessary to support critique. Notably, the contributions each demonstrate significant effort to develop communities of learners, within which guided discussion and shared experiences trust among members, developed through and through the valuing of participants' ideas, enables participants to freely and collaboratively explore otherwise difficult and sensitive issues around race and power (Vakil & Royston, 2019).

2.2. Humanizing data through creative representation

Another approach to fostering data literacy justice is in elevating emotion and embodiment through creative practices, which leverage multiple ways of knowing to communicate more effectively and equitably with data (D'Ignazio & Klein, 2020). In Kahn's et al. (2022) framework, Notice, Wonder, Feel, Act, and Reimagine, the authors highlight a procedure for integrating affect into one's approach to data. Three of the contributions in particular demonstrate how creating non-canonical representations engaged learners in new ways of sensemaking and communicating with data. For example, Bhargava et al. (2022) described how theatrical activities that allowed participants to embody statistical processes, such as binning (i.e., grouping data into categories), not only prompted learners' realization of how such processes can impact results, but also allowed both performers and audiences to relate to the stories of human experiences behind the data.

In Kahn et al. (2022) high school students, having calculated the minuscule chances of winning the lottery, used parody, analogies, and satire to communicate their findings, redesigning an existing lottery advertisement titled "Hey, You Never Know," to "Hey, Now You Know." The authors highlight how these rhetorical strategies were critical for learners to reimagine new futures for their community.

In Shapiro et al. (2022) data workers cleaned and explored a dataset on city policing activity through interactive visualizations, then created visual art to bring attention to aspects of their findings. Their design choices involved intentional decisions to humanize the data (e.g., plotting data points over a painted handprint, chosen to symbolize experiences in their own police encounters, such as in having fingerprints recorded during an arrest), what details to foreground (e.g., using black and red paint to distinguish the number of homicide-related arrests from other types of arrest); and what details to background (e.g., using aggregate neighborhood data to avoid drawing negative attention to particular neighborhoods).

As Philip Olivares-Pasillas and Rocha's (2016) study of visualization indicates, we cannot navigate authentic data without encountering issues of power that may reproduce and exemplify existing inequities. In each piece in this issue, these intersections are approached through creative visualization, bringing attention to the value of emphasizing alternative ways of representation, particularly those that center the contexts and people implicated in the data. Creative choices furthermore empowered learners to craft the messages they wanted to convey to the audiences they felt needed to receive them, thus giving agency to those directly impacted by the phenomena revealed in the data.

2.3. Writing the world with socially situated data artifacts

The learning experiences described in each of the contributions were notable in how they moved learners from critiquing data, to using data to take action toward just ends. One way that the contributions achieved this is through engaging audiences in sensemaking and advocacy, often by rethinking the relationships between form, medium and message.

For example, the data workers in Shapiro et al. (2022) created a publicly viewable video of their creative data visualizations of arrest data—juxtaposing their artwork with canonical data representations—to bring to light the disparate experiences between communities in their city, and to engage a broader audience in conversations around arrest data. The workshop participants in Bhargava et al. (2022) engaged in participatory theater, breaking down the typical barrier between performer and audience by inviting audience members to also perform, make meaning, and reflect alongside the performers. Similarly, Arastoopour et al. (2022) show how cooperative inquiry—a participatory design research approach that includes children in the design of their own learning experiences—with young learners facilitated them in rationalizing their choices to build algorithmic systems for particular purposes and users. Through their designs, children responded to the algorithmic bias embedded in many everyday technologies by embodying their own visions for more just applications of so-called helpful, data-driven technologies.

Together, these approaches demonstrate ways to support learners in acting to challenge the power structures with and within data practices; and in reimagining the structures, uses, and ownership of data (Khan et al., 2022). Resonant with the goals of culturally sustaining pedagogy, this approach goes beyond "How can 'we' get 'these' working-class kids of color to speak/write/be more like middle-class White ones" (Alim & Paris, 2017), to designing learning experiences that "perpetuate and foster—to sustain—linguistic, literate, and cultural pluralism as part of a schooling for positive social transformation" (Alim & Paris, 2017). At times such an orientation might involve participation in data work that emphasizes justice, and others, civic participation that draws on data knowledge. This perspective emphasizes that data-only approaches to exploring issues of justice can lead to exacerbation of inequities when data are used as the only or full story of an issue, or where data can be perceived to reify existing structures, rather than provide tools to challenge and construct structures through participatory action.

3. Directions for data literacy and social justice

Data literacy and justice are fundamentally interrelated, as each of the contributions to this special issue exemplifies. Further conceptual and empirical research is needed to conceptualize the key issues, tensions, and opportunities for learning at these intersections. Such insights will inform ways to support learning about, and toward, these interrelationships. Learners might bridge these domains by developing the cumulative knowledge and skills necessary for data justice: (1) Comprehension, that is, having an understanding the nature of data, and a fluency with manipulating it, (2) critique, which involves being aware of the sociocultural—and thus, non-neutral—nature of data, and noticing when data may represent, reify, and occlude issues of inequity and justice (D'Ignazio & Klein, 2020; Hardy et al., 2020), and (3) participation, which involves calling out and taking action on the biases inherent in the production and use of data, and advocating for social change. This special issue sits in a wider body of research on data literacy and justice both outside of, and within the field of education (particularly mathematics education). Here we draw together key lessons from this wider body of literature, and highlight areas for future development.

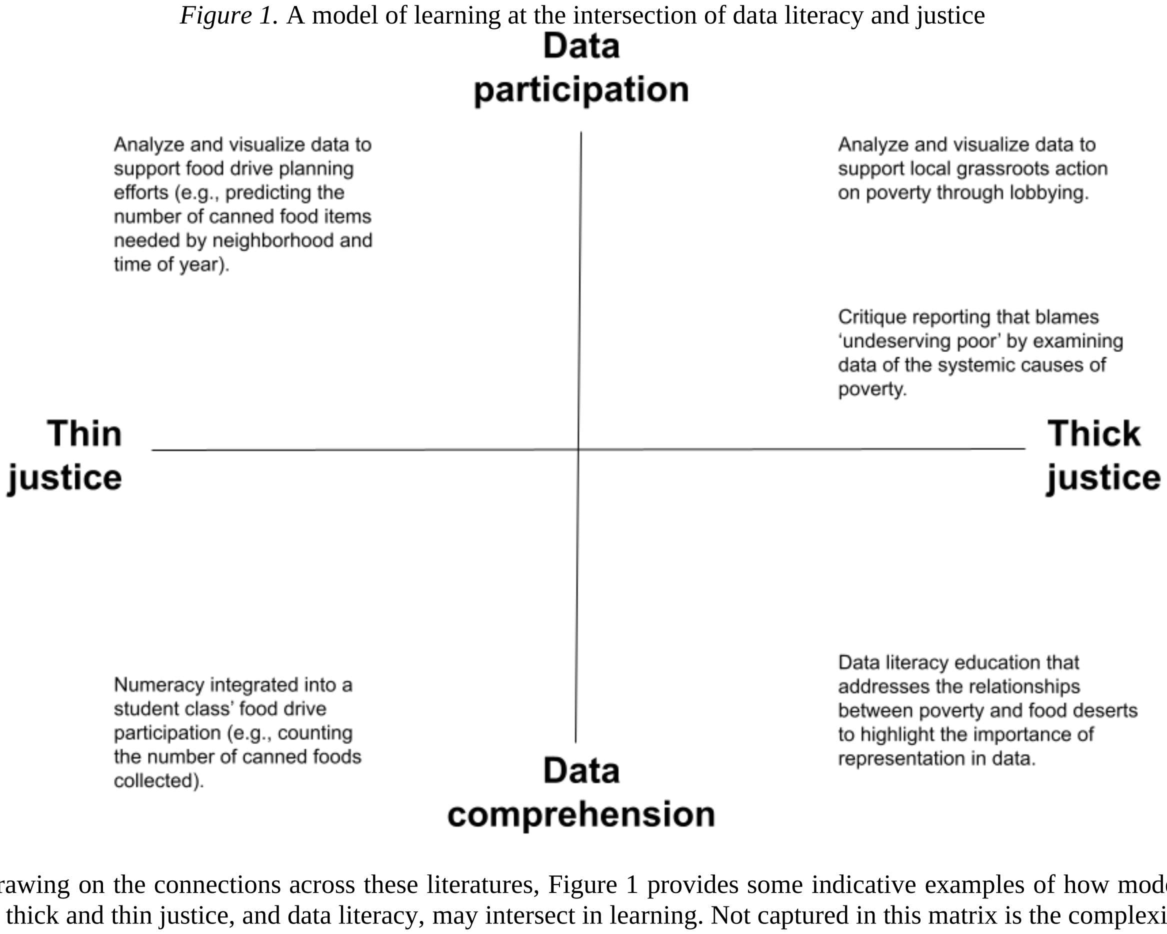
3.1. Thick and thin justice; data comprehension and data participation

Drawing on earlier work on thick and thin democratic education (e.g., Gandin & Apple, 2002; Carr, 2008; Zyngier, 2012), we suggest that one way to think about the intersection of data literacy and justice is through a thick and thin conception of justice. **Thin justice** foregrounds our participation as current or future members of justice or democratic activities and systems. This is what Westheimer and Kahne (2004) refer to as *personal responsibility models*, and which include paying taxes and obeying laws. **Thick justice** meanwhile seeks to promote participative action toward a deeper understanding of systems, not as given, but as shaped, shapeable, and critiqueable. This is what Westheimer and Kahne (2004) refer to as *justice-oriented models*, and which may include political organizing and lobbying.

In highlighting this spectrum, these researchers of democratic education emphasize that participation in society goes beyond knowledge of and participation in the structures "given," to shaping those structures through active democratic participation. This kind of participation is all the more important for tackling injustices and inequities reified in existing systems that can be challenging to tackle from within those systems. Through connecting thick justice to varieties of data literacy, we can highlight the potential to support learners in developing skills and knowledge such that through their comprehension, critique, and participation in and with data, they can write the world toward just ends.

In drawing this connection, we do not seek to make a value judgment regarding approaches taken; different approaches serve different purposes at different times for a range of audiences. Moreover, what might be described as relatively less critically engaged data literacy–what we describe as "comprehend," and has elsewhere been described in terms of reading the world–can serve to foster thick justice, in highlighting structures in society and how they change. In contrast, there are activities that involve data participation–the creation, critique, and engagement with data–that would not serve thick justice. This reflects the contextual nuance of these concerns, where features of the immediate micro (e.g., classroom, community center), meso (e.g., school, neighborhood), and macro (e.g., country, cross-nation), environment are likely to play important roles in interpreting actions and designing for justice learning.

Figure 1. A model of learning at the intersection of data literacy and justice



Drawing on the connections across these literatures, Figure 1 provides some indicative examples of how modes of thick and thin justice, and data literacy, may intersect in learning. Not captured in this matrix is the complexity of the content regarding key concepts, skills, and issues in both data and justice, ranging from the level of statistical or visualization skill, to the capacity to navigate nuanced legal or historical contexts. Navigation of this key knowledge is challenging, with a recent systematic review of empirical research indicating that teachers experience tensions in curricula design at the intersection of mathematics content and social justice (Xenofontos et al., 2021). Models to help understand, and navigate, these sociopolitical and learning contexts are needed.

As in wider development of citizenship oriented learning, there is a balance to be found between (1) delivering on the content of democratic and justice oriented learning, with the requisite need to understand the key concepts and issues learning is targeting, and (2) the risk of focusing on knowledge delivery, providing understanding of content that will be used in later democratic participation. Charting this path requires a rich pedagogical approach for justice citizens that reflects local contextual features and curricula (Heggart, 2021). Charting the path of justice, and its connection to data, similarly requires consideration of these issues. Do we have a clear idea of what students should learn about, with, and through enacting, justice concepts? And are we clear about the pedagogical tensions and challenges inherent in teaching toward learning at these intersections of data literacy and justice?

3.2. Charting a path toward data literacy justice

The studies described in this issue offer rich examples with which to begin conceptualizing the relationships between data literacy and social justice education. They also highlight the need to ensure meaningful engagement with both domains, particularly given the range and complexity of skills and concepts to be taught and understood, from statistics and visualization, to socio-political and historical ideas.

Together, the contributions raise questions to be considered as the field continues to explore learning at the intersections of data literacy and social justice. For example: How can critical approaches to data literacy and justice, such as arts-based methods, amplify the efforts of existing social activist communities (Bhargava et al., 2022). How can we support learners in reimagining possibilities for data technologies that challenge and surmount the bounds of mainstream perspectives and beliefs (Khan et al., 2022)? Relatedly, how, in reimagining the use of data and data-driven technologies for social good, can we build on the roles and expertise of non-professional members of society (Arastoopour Irgens et al., 2022)? How can criticality be embedded in data work to complement, rather than counter, the business-oriented priorities of professional workplace settings (Shapiro et al., 2022)?

Key issues come to light in considering the need for greater prominence of data literacy and social justice in the preparation of young learners. For instance, how can curriculum design for data literacy justice integrate into the existing disciplinary structures of formal educational institutions, and moreover, build upon and reflect the local contextual features of learners' social environments (Heggart, 2021)? How can we prepare educators to support learners in the interdisciplinary space between data literacy and justice (Xenofontos et al., 2021), particularly given the need for both domain content and pedagogical expertise relevant to both domains (Parker, 2018)?

Further clarity on the pedagogical tensions at the intersections of data literacy and justice will inform the development of necessary curricular resources and design guidelines, assessment and evaluation, educator development, and data technologies to support teaching and learning. Such work must address the needs of both learners and educators across settings including professional learning, K-16 classrooms, and informal learning contexts, particularly those engaged with civic participation. Through this work, the field may gain the necessary insight to move learners from data comprehension in service of thin justice goals, to data participation toward thick justice goals.

Acknowledgement

We thank the set of authors and reviewers who have contributed to this special issue. The work originated in two symposia at the International Conference of the Learning Sciences (ICLS) with a wider group of participants and contributors (Arastoopour Irgens et al., 2020; Matuk, et al., 2020), to whom we are grateful for those initial discussions. We are particularly grateful to Susan Yoon, Alyssa Wise, and Golnaz Arastoopour Irgens for their support in developing the issue proposal. We also thank Keith Heggart at UTS for helpful discussions regarding justice citizenship.

References

Alim, H. S., & Paris, D. (2017). What is culturally sustaining pedagogy and why does it matter? In D. Paris, & H. S. Alim (Eds.), *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world* (pp. 1-24). Teachers College Press.

American Statistical Association. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016*. http://www.amstat.org/education/gaise

Arastoopour Irgens, G., Knight, S., Wise, A. F., Philip, T., Olivares, M. C., Van Wart, S., Vakil, S., Marshall, J., Parikh, T., Lopez, L. M., Wilkerson, M. H., Gutiérrez, K., Jiang, S., & Kahn, J. B. (2020). Data literacies and social justice: Exploring critical data literacies through sociocultural Perspectives. In M. Gresalfi, & I. S. Horn (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020* (Vol. 1, pp. 406-413). International Society of the Learning Sciences. https://doi.org/10.22318/icls2020.406

Arastoopour Irgens, G., Adisa, I., Bailey, C., & Vega Quesada, H. (2022). Designing with and for youth: A Participatory design research approach for critical machine learning education. *Educational Technology & Society, 25*(4), 126-141.

Bhargava, R., Brea, A., Palacin, V., Perovich, L., & Hinson, J. (2022). Data theatre as an entry point to data literacy. *Educational Technology & Society, 25*(4), 93-108.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency* (pp. 77-91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1481

Carr, P. (2008). Educators and education for democracy: Moving beyond “thin” democracy. *Inter-American Journal of Education for Democracy*, *1*(2), Article 2. https://scholarworks.iu.edu/journals/index.php/ried/article/view/130

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics.* National Governors Association Center for Best Practices & Council of Chief State School Officers.

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. Information, *Communication & Society, 22*(7), 873-881. https://doi.org/10.1080/1369118X.2019.1606268

D’Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press. https://mitpress.mit.edu/9780262044004/data-feminism/

DiSalvo, B., & DesPortes, K. (2017). Participatory design for value-driven learning. In B. DiSalvo, J. Yip, E. Bonsignore, & C. DiSalvo (Eds.), *Participatory Design for Learning* (pp. 175-188). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315630830-17/participatory-design-value-driven-learning-betsy-disalvo-kayla-desportes

Druin, A. (2002). The Role of children in the design of new technology. *Behaviour and Information Technology*, *21*(1), 1-25.

Durrani, N., & Tariq, V. N. (2012). The Role of numeracy skills in graduate employability. *Education+ Training, 54*(5), 419-434. https://doi.org/10.1108/00400911211244704

Engel, J. (2017). Statistical literacy for active citizenship: A Call for data science education. *Statistics Education Research Journal, 16*(1), 44-49. https://iase-web.org/ojs/SERJ/article/view/213

FitzSimons, G., & Coben, D. (2009). Adult numeracy for work and life: Curriculum and teaching implications of recent research. In D. R. Maclean & D. Wilson (Eds.), *International Handbook of Education for the Changing World of Work* (pp. 2731-2745). https://doi.org/10.1007/978-1-4020-5281-1_179

Frankenstein, M. (1983). Critical mathematics education: An Application of Paulo Freire’s epistemology. *Journal of Education*, *165*(4), 315-339. https://doi.org/10.1177/002205748316500403

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report.* American Statistical Association.

Gal, I. (2002). Adults’ statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25. https://doi.org/10.2307/1403713

Gandin, L. A., & Apple, M. W. (2002). Thin versus thick democracy in education: Porto Alegre and the creation of alternatives to neo-liberalism. *International studies in Sociology of Education*, *12*(2), 99-116. https://doi.org/10.1080/09620210200200085

Gigerenzer, G. (2014). *Risk savvy: How to make good decisions*. Penguin Books.

Gutierrez, R. (2009). Embracing the inherent tensions in teaching mathematics from an equity stance. *Democracy & Education*, *18*(3), 9–16. https://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=44306819

Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education, 34*(1), 37-73. https://doi.org/10.2307/30034699

Gutstein, E. (2006). *Reading and writing the world with mathematics*. Routledge.

Hardy, L., Dixon, C., & Hsi, S. (2020). From data collectors to data producers: Shifting students' relationship to data. *Journal of the Learning Sciences, 29*(1), 104-126. https://doi.org/10.1080/10508406.2019.1678164

Heggart, K. (2021). *Activist citizenship education: A Framework for creating justice citizens*. Springer Nature.

Kahn, J. B., Peralta, L. M., Rubel, L. H., Lim, V. Y., Jiang, S., & Herbel-Eisenmann, B. (2022). Notice, wonder, feel, act, and reimagine as a path toward social justice in data science education. *Educational Technology & Society, 25*(4), 80-92.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kitchin, R. (2014). *The Data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Lee, O., & Campbell, T. (2020). What science and STEM teachers can learn from COVID-19: Harnessing data science and computer science through the convergence of multiple STEM subjects. *Journal of Science Teacher Education, 31*(8), 932-944. https://doi.org/10.1080/1046560X.2020.1814980

Levitin, D. J. (2016). *A Field guide to lies: Critical thinking in the information age*. Penguin Publishing Group.

Louie, J., Stiles, J., Fagan, E., Chance, B., & Roy, S. (2022). Building toward Critical Data Literacy with Investigations of Income Inequality. *Educational Technology & Society, 25*(4), 142-163.

Matuk, C., Yoon, S., Polman, J., Amato, A., Barton, J., Bulalacao, N. M., Cafaro, F., Haldar, L. C., Cottone, A., Cortes, K., DesPortes, K., Erickson, T., Finzer, W., Taylor, K. H., Herbel-Eisenmann, B., Graville, C., Gutiérrez, K., Higgins, T., Himes, B., Lanouette, K., Lee, H., Lim, V., Lopez, M. L., Lyons, L., Milz, D., Olivares, M. C., Osche, E., Parikh, T. S., Philip, T., Rubel, L., Shelley, J., Rivero, E., Roberts, J., Roberto, C., Petrosino, T., Rubin, A., Shim, J., Silander, M., Sommer, S., Stokes, D., Tes, M., Trajkova, M., Urbanowicz, R., Vacca, R., Van Wart, S., Vasudevan, V., Wilkerson, M., & Woods, P. J. (2020, June). Data literacy for social justice. In M. Gresalfi, & I. S. Horn, (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020* (Vol. 1, pp. 343-349). International Society of the Learning Sciences. https://doi.org/10.22318/icls2020.343

National Center for Science and Engineering Statistics (NCSES). (2021). *Women, minorities, and persons with disabilities in science and engineering*. http://hdl.voced.edu.au/10707/418545

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pangrazio, L., & Selwyn, N. (2019). "Personal data literacies": A Critical literacies approach to enhancing understandings of personal digital data. *New Media and Society, 21*(2), 419-437. https://doi.org/10.1177/1461444818799523

Parker, W. C. (2018). Human rights education's curriculum problem. *Human Rights Education Review, 1*(1), Article 1. https://doi.org/10.7577/hrer.2450

Pennington, D., Ebert-Uphoff, I., Freed, N., Martin, J., & Pierce, S. A. (2020). Bridging sustainability science, earth science, and data science through interdisciplinary education. *Sustainability Science, 15*(2), 647-661. https://doi.org/10.1007/s11625-019-00735-3

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A Framework for learning about big data with mobile technologies for democratic participation: Possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning, 18*(3), 103-120. https://doi.org/10.1007/s10758-013-9202-4

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction, 34*(4), 361-388. https://doi.org/10.1080/07370008.2016.1210418

PIAAC Numeracy Expert Group. (2009). PIAAC numeracy: A Conceptual framework [OECD Education Working Papers]. Organisation for Economic Co-operation and Development. http://www.oecd-ilibrary.org/content/workingpaper/220337421165

Quinn, D. M., & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher, 44*(6), 336-346. https://doi.org/10.3102/0013189X15598539

Raffaghelli, J. (2020). Is data literacy a catalyst of social justice? A Response from nine data literacy initiatives in higher education. *Education Sciences, 10*(9), 233. https://doi.org/10.3390/educsci10090233

Shapiro, B. R., Meng, A., Rothschild, A., Gilliam, S., Garrett, C., DiSalvo, C., & DiSalvo, B. (2022). "Bettering data": The Role of everyday language and visualization in critical novice data work. *Educational Technology & Society, 25*(4), 109-125.

Steen, L. A. (2004). *Achieving quantitative literacy: An Urgent challenge for higher education* (No. 62). The Mathematical Association of America.

Stone, D. (2018). The 2017 James Madison award lecture: The Ethics of counting. *PS: Political Science & Politics*, *51*(1), 7-16. https://doi.org/10.1017/S1049096517001767

Taylor, L. (2017). What is data justice? The Case for connecting digital rights and freedoms globally. *Big Data & Society, 4*(2). https://doi.org/10.1177/2053951717736335

Vakil, S., & Higgs, J. (2019). Education: It's about power. *Communications of the ACM*, *62*(3), 31–33. https://doi.org/10.1145/3306617

Vakil, S., & de Royston, M. M. (2019). Exploring politicized trust in a racially diverse computer science classroom. *Race Ethnicity and Education*, *22*(4), 545–567. https://doi.org/10.1080/13613324.2019.1592846

Van Wart, S. Lanouette, K., & Parikh, T. (2020). Third spaces for data science education using participatory digital mapping. *Journal of the Learning Science, 29*(1), 127-153, https://doi.org/10.1080/10508406.2019.1693378

Westheimer, J., & Kahne, J. (2004). What kind of citizen? The Politics of educating for democracy. *American Educational Research Journal*, *41*(2), 237–269. https://doi.org/10.3102/00028312041002237

Xenofontos, C., Fraser, S., Priestley, A., & Priestley, M. (2021). Mathematics teachers and social justice: A Systematic review of empirical studies. *Oxford Review of Education*, *47*(2), 135–151. https://doi.org/10.1080/03054985.2020.1807314

Yates, S. J., Carmi, E., Lockley, E., Wessels, B., & Pawluczuk, A. (2021). *Me and my big data: Developing citizens' data literacy*. Nuffield Foundation. https://www.nuffieldfoundation.org/project/me-and-my-big-data-developing-citizens-data-literacy

Kahn, J. B., Peralta, L. M., Rubel, L. H., Lim, V. Y., Jiang, S., & Herbel-Eisenmann, B. (2022). Notice, Wonder, Feel, Act, and Reimagine as a Path Toward Social Justice in Data Science Education. *Educational Technology & Society*, *25* (4), 80-92.

Notice, Wonder, Feel, Act, and Reimagine as a Path Toward Social Justice in Data Science Education

Jennifer B. Kahn[1*], Lee Melvin Peralta[2], Laurie H. Rubel[3], Vivian Y. Lim[4], Shiyan Jiang[5] and Beth Herbel-Eisenmann[2]

[1]University of Miami, USA // [2]Michigan State University, USA // [3]University of Haifa, Israel // [4]CUNY Guttman Community College, USA // [5]North Carolina State University, USA // jkahnthorne@miami.edu // peralt11@msu.edu // lrubel@edu.haifa.ac.il // vivian.liu@guttman.cuny.edu // sjiang24@ncsu.edu // bhe@msu.edu

[*]Corresponding author

ABSTRACT: In this paper, we introduce Notice, Wonder, Feel, Act, and Reimagine (NWFAR) to promote social justice in data science (DS) education. NWFAR draws on intersectional feminist DS to scaffold critical perspectives towards systems of power and oppression and attend to students' experiences in designs for learning. NWFAR adds three practices that are typically not emphasized in learning designs for DS: feel—engaging emotions and the physical body; act—challenging, inspiring, or informing others towards change; and reimagine—envisioning how data, data methods, and data technologies could pursue different problems, solutions, and perspectives. We illustrate NWFAR through two design-based research projects from prior empirical work. Through these two examples, we demonstrate what thinking with NWFAR could look like in practice and highlight future possibilities for learning. We conclude with a discussion that focuses on the reimagining dimension, in which we highlight social-justice oriented theories.

Keywords: Data science education, Data feminism, Critical data literacies, Social justice

1. Introduction

A data-driven world implies the essentiality of data practices (We prefer data practices to "data literacy," because we recognize there is a history of racialized and discriminatory uses of the term "literacy" (Philip & Rubel, 2019) or to data skills, because we view learning data science as a process of participation in situated practices with tools (Gutiérrez & Rogoff, 2003).). Accordingly, an enthusiasm for data science (DS) education—that is, education that leverages computer science, statistics, and mathematics knowledge for learning data practices—now extends to K-12 schools (e.g., LaMar & Boaler, 2021). The preK-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE II; Bargagliotti et al., 2020), for example, present "statistical literacy for all" as an "ultimate goal" (p. 5). However, stated goals like "statistical literacy for all" promote equity but with a generality that conceals structural inequities embedded within data practices that harm people from historically marginalized groups (Benjamin, 2019; Martin, 2003). Expanding access to data practices does not in and of itself respond to systemic inequities like racism or how the use of data might be a driver of injustice and oppression (Benjamin, 2019; Eubanks, 2018; Noble, 2018; O'Neil, 2016; Philip et al., 2013). Furthermore, the contemporary expansion of K-12 education to include DS, in terms of resources, technologies, and programs, rarely extends to education about how data and its uses could inspire or inform needed social change.

In this paper, we put forward a set of guiding questions to shape the design of learning environments for DS, focusing on learning environments that support students creating and/or interpreting data visualizations to inform decision-making and conclusions about real-world phenomena, with the goal of promoting social justice. We use Swalwell's (2013) conceptualization of social justice as one that:

> [R]ecognizes and affirms difference (e.g., cultural, sexual, political) while maintaining a commitment to fundamental human rights and democratic principles (e.g., freedom of speech and freedom of religion)...[C]hallenges the current distribution of resources in order to secure the basic needs required for human flourishing (e.g., safety, food, shelter, water, love)...[A]ssumes unequal power relationships and challenges the belief systems and social relations that (re)produce power differentials (p. 17, citing Nussbaum, 1992 and Fraser, 1997).

Social justice is an ongoing process, rather than an outcome, that must be continually sustained to attend to both material and ideological conditions that are always in flux. Further, we emphasize that this conceptualization of social justice does not locate sources of injustice within the minds and actions of biased individuals. Instead, a social justice orientation allows us to identify and challenge unequal power structures located in society at multiple scales.

We first discuss two current approaches to centering social justice in learning designs for DS. We then introduce and situate an instructional heuristic—Notice, Wonder, Feel, Act, and Reimagine (NWFAR; for earlier discussions of these ideas, see Rubel, 2020; Rubel et al., 2021a; Rubel et al., 2022)—in response to these approaches. We extend the "Notice and Wonder" instructional routine that originated with Burton (1984) by drawing on intersectional feminist perspectives on DS (D'Ignazio & Klein, 2020; Rubel et al., 2022). We then use NWFAR to reexamine two examples from previous empirical work from this authorship team. Finally, we point to scholarship outside of DS and statistics that offer new ideas for researchers and practitioners to work towards social justice goals.

2. Beyond humanistic and critical approaches to DS education

Existing guidelines for engaging students with data tend to emphasize technical skills and, beyond a few passing references to ethics, do not address its social uses and impacts. For example, GAISE II encourages educators to open opportunities for students to pose a question and then collect, interpret, and represent data towards answering their question (Bargagliotti et al., 2020). However, an approach to DS education that merely encompasses technical skills "leaves hardly any room for ethical and political considerations" (Bhargava et al., 2020), "disregards the need to address deeper structural issues of inequality" (Fotopoulou, 2021, p. 1641), and does not help students critique data, data practices, or the contexts in which data carry meaning (Philip et al., 2016). In the absence of adequate ethical and political considerations, individuals and institutions will continue to use data to support practices of oppression through technologies such as facial recognition software (Benjamin, 2019; Buolamwini & Gebru, 2018). For these reasons, it is imperative that educators promote the technical aspects of data practices alongside knowledge of and a desire to act with respect to issues of social justice and broader sociopolitical and cultural concerns related to data and data practices (Fotopoulou, 2021; Lee et al., 2021; Philip et al., 2016; Wilkerson & Polman, 2020; Stornaiuolo, 2020).

Two interrelated strands of research engage data practices with understandings of broader sociopolitical, cultural, and ethical concerns, the first of which takes a "humanistic stance" toward DS (Lee et al., 2021), while the second promotes "critical data literacies" (Tygel & Kirsch, 2016; Stornaiuolo, 2020; Fotopoulou, 2021). Lee et al. (2021) call for a humanistic stance that recognizes that data practices comprise intersecting personal, cultural, and sociopolitical layers that shape learners' experiences with data. They describe a humanistic stance as one that attends to students' personal and direct experiences with data, the cultural and sociotechnical infrastructures of data collection and use, and broader political and social narratives that affect data practices (We understand Lee et al.'s (2021) use of humanistic as different from "humanizing," such as in Paris & Winn's (2013) humanizing research, del Carmen Salazar's (2013) humanizing pedagogy, or Gutiérrez's (2018) rehumanizing mathematics, in how the latter uses of the term emphasize decolonization.). Lee at al. (2021) propose this broad humanistic stance as a starting place for more research and more specific frameworks that could support "critical participation" in data practices (p. 7), including the technical practices (i.e., generation of data, interpretation, argumentation) and considerations of how data is situated within systems of "power and privilege" (p. 4) typically associated with critical data literacies.

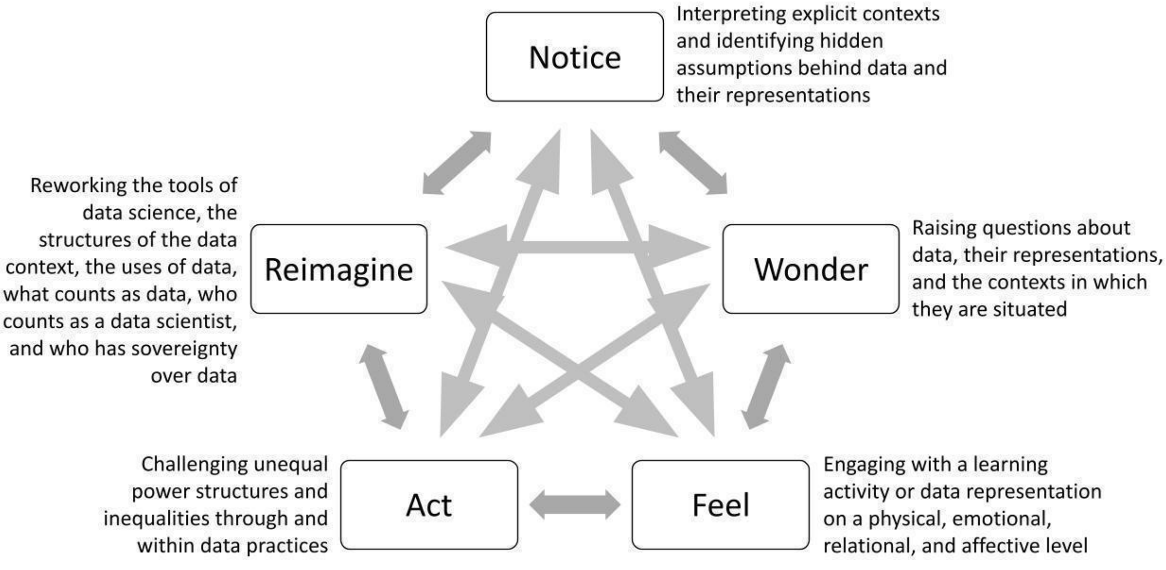
The strand of research that promotes critical data literacies often draws inspiration from Freire (1993). Tygel and Kirsch (2016), for example, frame critical data literacies in terms of reading, communicating, and producing data—all toward emancipatory goals. Others add considerations of when and where data are collected, how data are manipulated by hidden algorithms, and how ethics impact data practices, including concerns about privacy and the skewed incentive structures created by data practices (D'Ignazio & Bhargava, 2015; Hautea et al., 2017). Philip et al. (2016) argue that data practices must occur alongside deep interrogations of race and power, without which opportunities to learn about data can become counterproductive for failing to address connections between data practices and white supremacy. Stornaiuolo (2020) suggests that key objectives of critical data literacies include the opportunities for students to see themselves as agentic producers of data, to exploit data for their personal purposes, and to understand that data are "socially situated resources for meaning making" (p. 81; see also Pangrazio & Selwyn, 2019). Common themes across this body of work include the integration of technical data practices with sociopolitical awareness of power and identity as well as an orientation towards emancipation.

Common to both a humanistic stance toward DS and a critical data literacies approach is a focus on individual students' experiences with data, their agentic responses to data, and their capacities to intellectualize data and data practices with personal, technical, cultural, and sociopolitical lenses. Consequently, these perspectives may seem to overly emphasize the role of individual cognitive resources and rational responses to the way we experience the world (Bondi, 2005). That is, these perspectives treat rational processes of human decision-making and sensemaking as the primary objects of teaching and learning about data. However, as intersectional feminist scholars (e.g., D'Ignazio & Klein, 2020) and others interested in affect theories (e.g., Bondi, 2005) argue, aspects of engaging with the world, including aspects of our engagements with data and data practices in service of social justice, lie beyond the realm of rationalist explanation and sensemaking. These non-conscious and non-individual processes include feelings, as well as affective dimensions of social action and imagination, all of which are situated within relations of power among human and non-human individuals and collectives (D'Ignazio & Klein, 2020; Van Wijnendaele, 2011). Even justice-oriented researchers tend to privilege thinking and discourse over emotional, felt, and non-conscious experience as the primary means of acting with respect to social justice goals (Van Wijnendaele, 2011). In alignment with these critiques, we elevate feeling, action, and reimagination with thinking and discourse as data practices important to the process of social justice. Our proposed instructional heuristic, NWFAR, could be integrated into critical data literacies or humanistic approaches to DS. On one hand, to support critical data literacies, NWFAR encourages participation in analytic reasoning about data and its personal, cultural, and sociopolitical contexts. On the other hand, NWFAR also prompts teachers and students to engage with data affectively, by positioning feelings, actions, and the imagination as central aspects of DS education rather than peripheral ones, and to reflect on systems of power explicitly.

3. Notice, Wonder, Feel, Act, and Reimagine (NWFAR)

NWFAR functions as a kind of reflection tool for thinking about designs for learning from an intersectional feminist perspective. Such questions can inspire or influence new learning designs, for example, in the form of new or revised curricular materials, approaches to teaching, or forms of assessment, that consider issues of emotion, praxis, and the imagination, which are not captured by popular strategies such as the Notice and Wonder routine in use (Burton, 1984; Fetter, 2015; Gonchar & Schulten, 2017; National Council of Teachers of Mathematics, 2021). The "Notice and Wonder" routine begins when a teacher presents a data visualization and asks students "What do you notice?" and "What do you wonder?" The routine has multiple purposes, including facilitating participation, encouraging student sense-making, and communicating that there may be more than one correct answer to a problem. GAISE II, for instance, uses the language of noticing and wondering to prompt students to make observations about data (e.g., patterns or features of graphical displays, like axes or measures) and ask questions about the data pipeline (e.g., how, why, and for whom data were collected; Bargagliotti et al., 2020). NWFAR makes use of the instructional efficacy and popularity of Notice and Wonder in ways that foreground social justice.

Figure 1. Notice, Wonder, Feel, Act, Reimagine



NWFAR revises the "Notice and Wonder" routine in two ways. First, NWFAR reworks the notions of *notice* and *wonder* to be more closely tied to goals of social justice, rather than their original purposes of prompting questions for mathematical sensemaking. Second, NWFAR creates opportunities for learners to *feel*, *act*, and *reimagine* as responses to data and data practices, in addition to noticing and wondering. We draw on D'Ignazio and Klein's (2020) data feminism in developing these revisions. Data feminism is an intersectional feminist perspective that offers an alternative to the "male and white and technoheroic" paradigm of DS (D'Ignazio &

Klein, 2020, p. 9; see D'Ignazio & Klein (2020) for examples and elaborated discussion; Lunn et al., 2021), counters the prevalence of racism and sexism through DS, and calls for broader engagement of activists and community organizers in DS (e.g., Data for Black Lives, n.d.).

Our proposal of NWFAR is not intended to be a routine in the way that Notice and Wonder has been adopted among many mathematics educators, as a clearly defined ritual (see Rumack & Huinker, 2019). NWFAR is not a step-by-step approach to prompt students to recognize what they notice, wonder, feel, act, and reimagine in a particular order. Instead, it is a collection of possible starting points that can be approached and revisited in any sequence (Figure 1). In the following sections, we briefly expand on each of the elements of NWFAR and provide guiding questions for each dimension.

3.1. Notice

Notice refers to interpreting elements of the data visualization itself. These elements can include data sources, data trends, relationships between variables, comparisons between data points, identification of outliers as well as visual elements including titles, labels, annotations, or authorship information. Notice also refers to observing what is absent or left out. Questions that promote practices related to noticing but often are overlooked in typical curricular uses of notice include: (1) Whose perspectives are represented and whose are ignored? (2) How, when, and where was the data produced and by whom? (3) How does the visualization account for race, gender, class, and place and their intersection? (4) How do the authors describe the purpose of this visualization?

3.2. Wonder

Wonder means posing questions that are not directly answerable. Wonder raises questions about the data (and data that are missing), the process of producing and communicating data, and the contexts in which data are produced, for which there is no direct evidence in the data visualization. Questions that extend conventional prompts of "What do you wonder?" include: (1) What might the data visualization look like if it (more explicitly) acknowledged how race, gender, class, and place (and their intersections) play a role in the data context? (2) Who might benefit from the represented point of view, and who might profit from this dataset? (3) What might the data visualization look like if it more explicitly attended to human and non-human stakeholders and their physical bodies? (4) What might be some of the assumptions that the designers of the data visualization and activity make about people and relationships? (5) Who might have been involved in the collection, production, interpretation, analysis, visualization, and communication of the visualization?

3.3. Feel

Feel represents practices of engaging with a learning activity or data visualization on a physical, emotional, relational, and affective level. Embracing emotion represents an act of resistance against dominant practices in DS that tend to code reason as masculine and emotion as feminine and thus elevate reason but cast suspicion on emotion. Embracing emotion allows us to let go of the binary logic between reason and emotion and allow both to inspire action toward justice (D'Ignazio & Klein, 2020; e.g., hand-drawn data illustrations by Chalabi (Rakotondravony, 2019)). Through an emphasis on feeling, we follow the "affective turn" in feminist theory to challenge the notion of universal objectivity as a privileged goal when engaging with data (D'Ignazio & Klein, 2020; Pedwell & Whitehead, 2012). Data practices can and should include visceral experiences that involve the body, physically and emotionally (Malinverni & Pares, 2014). This includes engaging the body through senses such as touch and the feelings that these senses generate (Lupton, 2017), through embodied practices such as gestures and movement to interact with and represent data (Roberts & Lyons, 2019), and through emotions (Kennedy & Hill, 2018). Questions include: (1) How does this data visualization engage your body, your senses, and your emotions? (2) How might this visualization make space for multiple kinds of feelings? (3) How does the data or context connect to interpersonal relationships or to relationships with nonhuman others and the planet?

3.4. Act

Act refers to challenging unequal power structures and exposing inequalities through and within data practices. D'Ignazio and Klein (2020) offer examples for what acting can look like in DS: compiling counterdata, challenging power by analyzing inequitable outcomes across different groups, working towards co-liberation, and cultivating a new generation of data feminists. As we argued in earlier work (Rubel et al., 2021b), action directed toward transforming systems of power must go beyond merely "playing the game" of DS. Instead, the act dimension of NWFAR must also include seeking to "change the game" through critique and social action (Rubel et al., 2021b; see Gutiérrez, 2011). Questions include: (1) How can you use the data or data visualization to challenge and inform, inspire, or support political action? (2) How does this data visualization raise ethical questions, including questions about data security and data privacy? (3) What might you learn from reaching out to others about information presented in this data visualization? (4) In what ways are you moved to act, to refrain from action, to make change, or to communicate to others?

3.5. Reimagine

Reimagine describes practices directed toward reworking DS tools, as well as the structures of the data context, its uses, what counts as data, who counts as a data scientist, and who has sovereignty over data. For instance, reimagining could entail thinking about how a data visualization could produce a different solution or show different perspectives. This use of reimagination draws on the data feminism principle of elevating emotion and embodiment in DS (D'Ignazio & Klein, 2020; Ioannou & Ioannou, 2020) by challenging (e.g., artist Onuoha's (2016) Library of Missing Datasets calling attention to the absence of data) and reinventing existing data visualizations and data stories. For example, the Data Zetu program in Tanzania created a fashion competition and show in which fashion designers reimagined how open public health data could be used by designing clothing for women that could spark conversations around sexual and reproductive health and sexual and gender-based violence (Katuli, 2018).

Warren (2021) offers a way of thinking about reimagining, drawing on Kelley's (2002) *Freedom Dreams: The Black Radical Imagination*. Warren's framework for centering possibility in Black education comprises nine interlinked concepts: Resistance, Dreaming, Storytelling, Creativity, Thriving, Community, Reparations, Environment, and Teaching. We expand on resistance (what needs to be torn down) and dreaming ("what should be built atop the rubble," p. 31) because of their importance in Warren's teachings about a future-oriented Black education transformation and their potential for developing designs that make reimagining a more central aspect of learning with data. Resistance "encompasses multiple forms of opposition to education conditions, broadly conceived, that insist on Black folks' subordination to Eurocentric expectations for thinking, speaking, and being" (Warren, 2021, p. 21). Resistance requires critical reflection on how schooling, and, for purposes of this paper, data and their contexts, necessitate opposition to unjust authority figures and systems. Dreaming provides a means for reimagining the capabilities of Black (and other marginalized) individuals to "produce novel creations that respond to or solve a problem of interest to the maker: designs guided primarily by the desire of the producer versus that of the colonizers" (Warren, 2021, p. 36). Warren's articulation of resistance, dreaming, and creativity link to the kinds of speculative fiction, science fiction storytelling, counter-storytelling, and allegorical storytelling that have begun to appear in STEM education work. In the discussion following our example analysis, we point to theoretical perspectives that offer more radical approaches to reimagining that could move DS education further towards social justice and equity goals.

4. Returning to prior work to illustrate and extend NWFAR

In this section, we revisit prior work from among this authorship team using NWFAR. Both projects prompted youth to explore questions about socioeconomic issues that resonated with their personal, family, or community experiences through activities that included exploring georeferenced, large-scale datasets and digital data visualization interfaces. The first example, the Lottery Project, was a design-based project, in collaboration with a mathematics teacher, that engaged high school students in investigating the role and impact of the lottery on the local community (Rubel et al., 2016a; Rubel et al., 2016b; Rubel et al., 2017). After learning how to calculate the probability of winning various lottery games, the students collected data about people's experiences with the lottery and analyzed lottery spending across neighborhoods. The second example, the Family Migration Project, comes from a design-based research study that asked teenage youth to investigate personal family migration histories with data in a summer program (Kahn, 2020). Students explored family histories and experiences with family members and connected those to analyses of socioeconomic data trends to tell the stories of what moved their families.

We use NWFAR as a tool to analyze the intended and enacted designs. More specifically, we point to the projects to demonstrate the feeling and acting components of NWFAR. We look across these two examples to identify learning opportunities that could have been supported if one were to redesign with NWFAR in mind. We show that our examples of reimagining are limited; when prompted, students' reimagining stayed within existing structures or systems. In our discussion, we point to possible directions for future iterations of NWFAR.

4.1. Feeling and acting in the Lottery Project

The Lottery Project is a curricular unit engaging high-school students to learn about the lottery as a public system and its impact on their community (Rubel et al., 2016a; Rubel et al., 2016b; Rubel et al., 2017). Students learn about the lottery by calculating the probability of winning local lottery games, analyzing ticket sales data across their city, and noting how the state utilizes lottery revenue. They learn about the impact of the lottery in their community by interviewing people about their lottery participation and mapping where lottery tickets are sold. In retrospectively revisiting the learning designs for the Lottery Project using NWFAR, we identify attempts in the design to engage students to feel by drawing on their own and their community members' lived experiences to inform critical analyses of the lottery system and prompt students to act and reimagine in response. All descriptions below are based on field notes and artifacts from classroom observations conducted in a high-school class in a large city in the Northeastern United States. Artifacts included photographs of student work and of the classroom whiteboard.

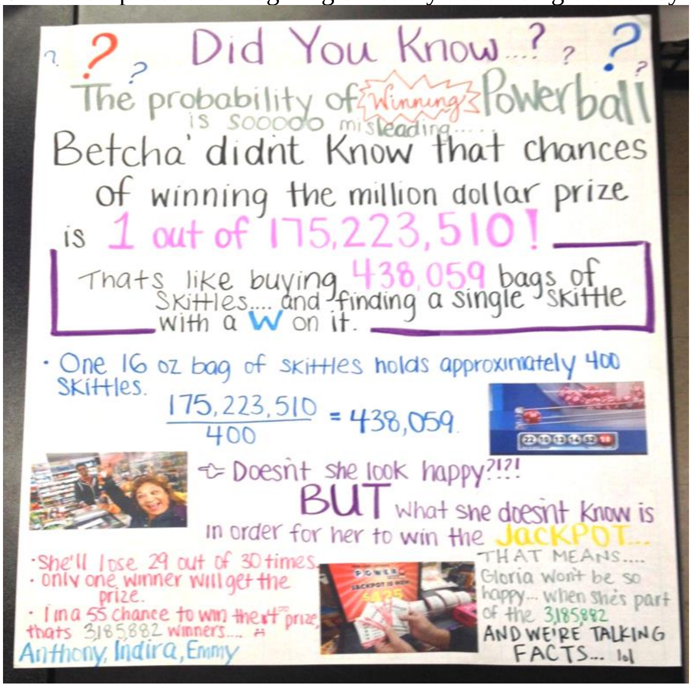
Various activities in the Lottery Project prompted students to feel. For example, a lesson early in the unit asked students to contribute their free associations with the lottery to create a collective Word Wall. The open prompt engaged students' personal reactions. Through the Word Wall activity, we learned that some students sensed that the lottery is unfair, for example, with some expressing their chagrin that "Black people don't win." These collective ideas would factor into students' stance toward questioning the lottery system. Through the project, there were reflection opportunities, in which students could express other feelings. Examples of expressions of feeling included surprise about the slim chances of winning and that this system is predesigned and owned by the state; outrage about the lottery operating as a "scam" targeting low-income areas; resignation about the lottery operating under the notion of free choice; and promise about the lottery providing hope and funding for education.

The Lottery Project included a field research component, in which students extended their inquiry into the lottery's presence and impact into the local neighborhood. This field research engaged participants' physical bodies by their interaction with people, places, and their physical environments. By interviewing people in the neighborhood, students could supplement their own experiences and add complexity to the personal stories about the relationships that people have to the lottery. For example, in a class debrief of their findings from these interviews, students noted that some people buy lottery tickets out of an addiction, but many view the lottery as a chance to satisfy basic needs and change their lives—paying children's college tuition, providing housing for family members, or returning to their home countries. The field research enabled students to consider personal and emotional elements in their understanding of the lottery alongside learning about the low probability of winning these lottery games. They learned that playing the lottery is not necessarily indicative of poor mathematical decision-making but rather an almost guaranteed outcome of an institution that is designed to prey on the most vulnerable people.

The project's technology enabled multilayered analyses of qualitative information about peoples' experiences in context and quantitative data about lottery spending and revenues. Students explored interactive maps showing daily lottery spending, median household income, and the ratio between these two figures by neighborhood. The lessons guided students to make comparisons among neighborhoods and notice and wonder about how the ratio of lottery spending to income is higher in lower-income neighborhoods. These interactive data visualizations, created by the research team, were the base layers for embodied, qualitative data that the students gathered. That is, students captured audio recordings, photos, and text from neighborhood interviews and uploaded these data as geolocated objects. Thus, students connected and grounded their analyses of the city-scale quantitative data to the local-scale and vice-versa. As a result, students had the opportunity to notice, wonder, and feel the lottery's impact in a more nuanced way.

The project also created opportunities for acting and reimagining. For example, using combinatorics to calculate the miniscule (yet hidden) probability of winning various state lottery games, students created posters to speak back to the state's "Hey, You Never Know" advertising campaign. Through these "Hey, Now You Know" posters, students reimagined the public lottery messaging into educative forms. For example, a group of students likened the chances of winning the Powerball game to finding a single, marked Skittle among 438,059 bags of skittles (Figure 2). The students included a photo of a gleeful-looking woman posing with lottery tickets in hand, captioned, "Doesn't she look happy?!?!" followed by a "BUT" and a list of caveats based on probability calculations, concluding that "Gloria won't be so happy" when she realizes she is part of the large number of losers. A second set of activities prompted students to reimagine the lottery as a system. As this played out, students primarily considered how the lottery could be redesigned to produce more winners from their community. Students were hesitant, despite guidance, to fundamentally reimagine the lottery system, as a source of public funding, private income, state revenue, or as a way for people to change their lives. We interpret this hesitation as shaped by how infrequently social justice is engaged in education as a matter of reimagining basic systems that are indeed transformable.

Figure 2. An example of a reimagining of lottery advertising created by students



4.2. Feeling and acting in the Family Migration Project

A second project (Kahn, 2020) offers an example that moved beyond the traditional notice and wonder pedagogy by centering family relationships and conversations as a starting point for data exploration. In this project, students (n = 17, mostly African American, ages 11-16) participated in a free summer program at a public library in a midsized southern city. Students connected their family migration histories (where they live, where they moved to), or *family geobiographies*, to the socioeconomic push and pull factors that drive family migration through an exploration of demographic datasets using two data visualization tools (Social Explorer, Gapminder). All descriptions below were based on interaction analysis of video and audio records of activities, including screen capture recordings of data exploration on laptops, as well as fieldnotes and students' final PowerPoint family data stories.

By starting with a retrospective, personal inquiry, we created spaces for the NWFAR dimensions of feeling and acting by providing an opportunity for families to be involved in learning. The Family Migration project began by asking students to trace their family geobiography back to their grandparents' generation. The family geobiography positioned the perspective of families at the center of the exploration of the aggregate data. During the workshop, the project leveraged students' relationships with siblings also participating in the workshop and other family members. Students were encouraged to text or call parents or relatives during the sessions and talk with their family on the intervening nights between workshop days. Family members consequently participated in the assembling of data stories, and students worked to resolve and/or represent multiple voices and stakeholders. This distributed effort (Kahn, 2020) challenged the traditional DS approach to human individuals as isolated data points. Similarly, in informal interviews at the end of the project, nearly all students said they would talk to family members to answer the questions that their projects raised. We view these conversations with family members, both those that occurred in the workshop and those intended for the future, as related to acting. The involvement of families and the telling of family histories also opened a space for feeling and emotion while engaging with the data.

The design incorporated traditional approaches to notice and wonder around interpreting data displays. We asked students to notice and wonder about aggregate-level trends in the data models and maps. However, we consistently situated noticing and wondering practices in familial and personal contexts. For example, one of the first activities asked students to brainstorm (wonder) about what moved families in general and then explore those factors with variables using the data tools. Students looked for differences across places (where their family member lived previously compared to where they moved to) on variables like income and education. Variables were sometimes selected by students on their own, other times suggested by instructors. Once students selected variables, instructors encouraged them to describe visual trends in the data. The interactivity of the data visualization web-based tools and the breadth of available data sets supported both practices.

The personal nature of noticing and wondering supported enactments of feeling and acting. For example, students frequently noticed that data that represented their family members' experiences was unavailable or that the data did not align with the story they wanted to tell. In turn, they pursued different strategies, such as the inclusion of multiple data displays in their data stories, to resolve this trouble in the data modeling environments (Jiang & Kahn, 2020). Steps to address the conflict could also be viewed as a form of acting with the data tools.

However, sometimes feelings clashed with noticing and wondering. For example, several students wanted to pursue stories about why their family moved for which they could not find a related dataset. For instance, one student wanted to tell the story of how her mother moved from Thailand to the United States for love—to marry her father. Love, however, is a variable that does not exist in traditional demographic datasets. Another student wanted to tell a data story about her family's celebration of Kwanzaa but similarly could not find an appropriate demographic variable to represent the population that celebrates this holiday. In future iterations, these conflicts between family stories and the limitation of datasets could be leveraged as opportunities for reimagining the data and data visualization.

The project incorporated several embodied data activities as well that were intended to serve as resources for reasoning about historical socioeconomic conditions from a relational perspective and with affect, although these were outside of the data visualization environment (as opposed to the exploration of data displays driven by body movements as in Roberts & Lyons, 2020). Students performed a walking-scale timeline that asked them to stand in for the family in historical time; we repeated this embodied activity with parents and students at the culminating public showing of student work. In all sessions, we also asked students to participate in a four walls game (like a four corners game). When asking the students to select a family member, we were also asking them to physically stand in and speak for their family members, who potentially had survived or witnessed discrimination and prejudice. For instance, questions about occupations raised additional queries like what constitutes "work," particularly for students thinking of enslaved ancestors, with some students stating that any forced labor under slavery should not qualify as an occupation. We also wanted students to be able to assemble stories or models grounded in experiences of people of color and challenge dominant social ideologies of equality and opportunity if they wanted to tell these stories. We found that students recognized their families' struggles in their data stories but focused on the positive outcomes for their family members. Being in charge of their own stories, including what data to show, was powerful for students.

Finally, students were not asked to reimagine the data or data tools in this project. In this way, the use of existing data visualization platforms and datasets arguably restricted the enactment of this part of NWFAR. However, our design did have students' futures in mind (e.g., we asked students where they see themselves living when they grow up). Additionally, some students considered what their lives might have been like if their family had not moved and assembled data stories around this reimagination of their own lives (a hypothetical counterfactual; Kahn, 2020). Below we discuss other possibilities for reimagination.

5. Discussion

NWFAR draws on intersectional feminism to examine power and promote equity within DS. The values and ethics embedded in this instructional heuristic depart from existing approaches to DS education. DS tends to be a white male space that emphasizes individuality, competitiveness, and the development of technical and cognitive skills but leaves out other ways of knowing and being. With NWFAR, we seek to add a new perspective to the growing body of scholarship committed to countering the creation and perpetuation of harmful and discriminatory data-driven practices that contribute to the marginalization of women, people of color, and other historically marginalized groups.

Our reflections on two projects supported the development of NWFAR. The two projects are examples that illustrate how learning designs can make space for feelings, both physically and emotionally, and motivate or inspire action in ways that are productive for developing critical perspectives towards issues of power, equity, and social justice. Both examples implemented notice and wonder in ways that attended to personal or local community histories and experiences and emphasized feelings and emotions in data engagements. Students were guided to act in various ways, such as creating posters as part of a public messaging campaign and having conversations with family members about their geobiographical histories.

There were moments in these projects when students engaged in practices of reimagining, but broader reworkings of the tools, structures, and uses of DS were not incorporated into the designs. Students were asked to reimagine public messages around the lottery (e.g., *Hey, Now You Know* posters) or reimagine the lottery system itself. In the family migration project, some explored what their lives would have been like if their families had a different geobiographical story. There is ample room to deepen how we might engage students in reimagining. In the Lottery Project, for example, we could have asked students to generate ideas for creating alternative opportunities for investment that would directly benefit their own communities. More fundamentally, we could have prompted students to consider what it would take for society to ensure that no one could be vulnerable to relying on the lottery for basic needs and resources. In the family migration project, we could have asked students to reimagine data that included their family members' experiences or to develop their own alternative history narratives, incorporating details about how their lives and the lives of their families and communities might have shaped and been shaped by broader sociopolitical and historical forces. In this way, reimagining becomes a way to use ideas about desirable futures and alternative pasts as tools to aid in speculative thought about how life could or should be in the present. Thus, reimagining practices seek to center justice, dignity, and freedom, especially for marginalized peoples and communities.

Additionally, our projects used public DS resources (open data visualization applications) as pedagogical tools to foster students' data practices. However, the use of professional data technologies in instructional settings raises questions for how to design and use these resources to support critical data practices as well as relational, affective learning experiences. In our project examples, to accomplish this, we intentionally connected activities outside the data interfaces that generated qualitative data (e.g., brainstorms, field research, walking-scale timelines) to data visualizations. Alternatively, there are other possibilities for achieving the goals of NWFAR within the technologies themselves. For example, while new DSE research has looked at developing tools for students to build machine learning models with data to gain a fundamental understanding of artificial intelligence (AI; e.g., Biehler & Fleischer, 2021), the activity and AI interface could be designed to help students learn the subjectivity of model development and decisions made by AI, and how AI can perpetuate biases and unfairness. Both data technologies and the learning experiences created around them should be redesigned to support complex reasoning about data to help students to engage with issues related to power, equity, and social justice.

5.1. Future directions: Developing "reimagine" in NWFAR

The act of reimagining—oneself, one's community, and one's society—is a powerful but often overlooked data practice. As we have reflected, the projects could have more deeply engaged students in practices of reimagining to better support social justice and equity. In line with how data feminism relies on broad and diverse examples and intellectual traditions, we turn to literature outside of DS education, including Black feminist thought and Indigenous quantitative methodologies. These theories draw on various frameworks and intellectual traditions, including critical Afropessimism (as opposed to vulgar Afropessimism, see Woodson, 2021), radical hope (Grant; 2021; Kelley, 2002), Black liberatory fantasy (Dumas & Ross, 2016), critical race theory (Crenshaw, 1991; Davis & Jett, 2019), BlackCrit (Dumas & Ross, 2016; Martin et al., 2019), Afrofuturisms (Alexander, 2019; Bell, 1992; McGee & White, 2021), Indigenous Futurisms (Dillon, 2012) and alternative histories (Carroll, 2020; Rick, 2021). The frameworks are necessarily interdisciplinary, allowing for new connections across identities and disciplines (McGee & White, 2021).

Speculative education critiques historical and inequitable schooling conditions and seeks to actively reimagine and create new possibilities for education (Mirra & Garcia, 2020). Mirra and Garcia (2020) worked with classroom communities across the United States, asking students to define their civic communities, engage in conversation with peers around a chosen civic topic, and imagine and represent potential civic futures. Their third design cycle included opportunities for students to craft origin stories for their own "civic superheroes" and explore how such superheroes could address issues they found to be pressing. Mirra and Garcia (2020) found that the processes of imagination did not come easily to youth or adults as efforts of imagination came up against the reality of clashing political ideologies among students. Thus, opportunities that include supporting students to craft speculative narratives based on real-world datasets likely need attention to how practices of imagination might confront entrenched political beliefs or dominant perspectives about the purposes and possibilities of data and data technologies.

We emphasize that reimagining could become an empty gesture without careful consideration of the contexts in which the research or design work is being done. We recognize that ideas about speculative education, Afro- and Indigenous futurisms, and radical hope, among others, need to be thoughtfully and deeply applied to not overlook oppressive histories and further damage or harm historically marginalized populations. Using these theoretical perspectives requires serious consideration of whiteness, white supremacy, antiBlackness and patriarchy while engaging with these liberatory ideas, a task that may be challenging as DS and technology design fields continue to be dominated by white men.

6. Conclusion

NWFAR encourages DS education researchers and practitioners to raise important questions about data and data visualizations and look for opportunities to explore emotion, praxis, and imagination. Our conceptualizations of noticing, wondering, feeling, acting, and reimagining seek to go beyond mere participation in or access to DS and move toward embracing new forms of engagement with data and data practices, including an emphasis on affect and the relationship between data and issues of power. In this sense, NWFAR is a heuristic that directs DS education towards social justice. The next steps are to implement and study NWFAR in various contexts, including learning designs for K-12 education and teacher education. This process will generate information about how to revise the heuristic and refine its set of reflective questions.

Acknowledgement

This material is based upon work done while Beth Herbel-Eisenmann is on assignment at the National Science Foundation. This material is based on work supported by the National Science Foundation under Grant No. 1341882 and Grant No. DRL-1222430. Any opinions, findings, and conclusions or recommendations expressed in this material are our own and do not necessarily reflect the views of the National Science Foundation.

This article is based on a conference proposal presented at the 2021 Annual Meeting of the American Educational Research Association:

Peralta, L. M., Rubel, L., Herbel-Eisenmann, B., Lim, V., Kahn, J., & Jiang, S. (April, 2021). What's going on in "What's going on in this graph?" Research paper presented at the Annual Meeting of American Educational Research Association, Online.

References

Alexander, N. N. (2019). Daija's awakening: CRT and Afrofuturism in mathematics education. In J. Davis & C. C. Jett (Eds.), *Critical race theory in mathematics education* (pp. 56-74). Routledge.

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *PreK-12 guidelines for assessment and instruction in statistics education ii (GAISE II): A framework for statistics and DS education*. NCTM & ASA.

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). Beyond data literacy: Reinventing community engagement and empowerment in the age of data. *Data-Pop Alliance White Paper Series.* Harvard Humanitarian Initiative, MIT Lad and Overseas Development Institute Internews. https://dspace.mit.edu/bitstream/handle/1721.1/123471/Beyond%20Data%20Literacy%202015.pdf

Bell, D. (1992). *Faces at the bottom of the well: The permanence of racism*. Basic Books.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43, S133-S142.

Bondi, L. (2005). Making connections and thinking through emotions: Between geography and psychotherapy. *Transactions of the Institute of British Geographers*, *30*(4), 433-448.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research)* (Vol. 81, pp. 77-91). PMLR.

Burton, L. (1984). Mathematical thinking: The Struggle for meaning. *Journal for Research in Mathematics Education*, *15*(1), 35-49.

Carroll, A. (2020). *The Man in black: Alternate history where superheroes try to save the world* [eBook edition]. https://www.amazon.com/Man-Black-Alternate-History-Superheroes-ebook/dp/B0893NNXLW

Crenshaw, K. W. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*, 1241-1299.

Data for Black Lives. (n.d.). *Data for Black Lives is a movement of activists, organizers, and scientists committed to the mission of using data to create concrete and measurable change in the lives of Black people.* http://d4bl.org/

Davis, J. & Jett, C. C. (Eds.). (2019). *Critical race theory in mathematics education.* Routledge.

del Carmen Salazar, M. (2013). A Humanizing pedagogy: Reinventing the principles and practice of education as a journey toward liberation. *Review of Research in Education*, *37*(1), 121-148.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to building big data literacy. *Proceedings of the Bloomberg Data for Good Exchange 2015*. http://www.kanarinka.com/wp-content/uploads/2015/ 07/Big_Data_Literacy.pdf

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.

Dillon, G. L. (Ed.). (2012). *Walking the clouds: An Anthology of indigenous science fiction.* University of Arizona Press.

Dumas, M. J., & Ross, K. M. (2016). "Be real Black for me": Imagining BlackCrit in education. *Urban Education*, *51*(4), 415-442.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Fetter, A. (2015, June 8). *Ever wonder what they'd notice?: Annie Fetter* [Video]. Youtube.com. https://www.youtube.com/watch?v=a-Fth6sOaRA

Fotopoulou, A. (2021). Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information, Communication & Society*, *24*(11), 1640-1657.

Fraser, N. (1997). *Justice interrupts: Critical reflections on the "postsocialist" condition.* Routledge.

Freire, P. (1993). *Pedagogy of the oppressed* (Rev. ed.). Continuum.

Gonchar, M., & Schulten, K. (2017, September 6). Announcing a new monthly feature: What's going on this graph? *The New York Times.* https://www.nytimes.com/2017/09/06/learning/announcing-a-new-monthly-feature-whats-going-on-in-this-graph.html

Grant, C. A. (2021). Radical hope, education and humanity. In C. A. Grant, A. N. Woodson, & M. J. Dumas (Eds.), *The future is Black: Afropessimism, fugitivity, and radical hope in education* (pp. 65-71). Routledge.

Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, *32*(5), 19-25.

Gutiérrez, R. (2011). Context matters: How should we conceptualize equity in mathematics education? In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in discourse for mathematics education: Theories, practices, and policies* (pp. 17-34). Springer.

Gutiérrez, R. (2018). Introduction: The Need to rehumanize mathematics. In I. M. Goffney & R. Gutiérrez (Eds.), *Rehumanizing mathematics for Black, Indigenous, and Latinx students*. *Annual Perspectives in Mathematics Education.* National Council of Teachers of Mathematics.

Hautea, S., Dasgupta, S., & Hill, B. M. (2017). Youth perspectives on critical data literacies. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 919-930). http://doi.org/10.1145/3025453.3025823

Ioannou, M., & Ioannou, A. (2020). Technology-enhanced embodied learning. *Educational Technology & Society*, *23*(3), 81-94.

Jiang, S., & Kahn, J. (2020). Data wrangling practices and collaborative interactions with aggregated data. *International Journal of Computer-Supported Collaborative Learning*, 15(3), 257-281.

Kahn, J. (2020). Learning at the intersection of self and society: The Family geobiography as a context for data science education. *Journal of the Learning Sciences*, *29*(1), 57-80.

Katuli, M. (2018, March 28). Young artists use fashion and data to promote dialog on sexual health. *Medium*. https://medium.com/data-zetu/young-artists-use-fashion-and-data-to-promote-dialog-on-sexual-health-517429662ec2

Kelley, R. D. (2002). *Freedom dreams: The Black radical imagination*. Beacon Press.

Kennedy, H. & Hill, R. L. (2018). The Feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology*, *52*(4), 830-848.

LaMar, T., & Boaler, J. (2021). The Importance and emergence of K-12 DS. *Phi Delta Kappan*, *103*(1), 49-53.

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A Call for a humanistic stance toward k–12 DS education. *Educational Researcher*, *50*(9), 664-672.

Lupton, D. (2017). Feeling your data: Touch and making sense of personal digital data. *New Media & Society*, *19*(10), 1599-1614.

Lunn, S., Zahedi, L., Ross, M., & Ohland, M. (2021). Exploration of intersectionality and computer science demographics: Understanding the historical context of shifts in participation. *ACM Transactions on Computing Education (TOCE)*, *21*(2), 1-30.

Malinverni, L., & Pares, N. (2014). Learning of abstract concepts through full-body interaction: A Systematic review. *Educational Technology & Society*, *17*(4), 100-116.

Martin, D. B. (2003). Hidden assumptions and unaddressed questions in mathematics for all rhetoric. *The Mathematics Educator*, *13*(2), 7–21.

Martin, D. B., Price, P. G., & Moore, R. (2019). Refusing systemic violence against Black children: Toward a Black liberatory mathematics education. In J. Davis & C. C. Jett (Eds.), *Critical race theory in mathematics education* (pp. 32-55). Routledge.

McGee, E. O., & White, D. T. (2021). Afrofuturism: Reimaging STEM for Black urban learners. In H. R. Milner & K. Lomotey (Eds.), *Handbook of Urban Education* (pp. 384-396). Routledge.

Mirra, N., & Garcia, A. (2020). "I hesitate but I do have hope": Youth speculative civic literacies for troubled times. *Harvard Educational Review*, *90*(2), 295-321.

National Council of Teachers of Mathematics. (2021). *What is notice and wonder?* https://www.nctm.org/noticeandwonder/

Noble, S. U. (2018). *Algorithms of oppression: Data discrimination in the age of Google*. NYU Press.

Nussbaum, M. (1992). Human functioning and social justice: In defense of Aristotelian essentialism. *Political Theory*, *20*(2), 202-246.

Onuoha, M. (2016). *The Library of missing datasets* [Mixed-media installation]. https://mimionuoha.com/the-library-of-missing-datasets

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Pangrazio, L., & Selwyn, N. (2019). "Personal data literacies": A Critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419-437.

Paris, D., & Winn, M. T. (Eds.). (2013). *Humanizing research: Decolonizing qualitative inquiry with youth and communities.* Sage Publications.

Pedwell, C., & Whitehead, A. (2012). Affecting feminism: Questions of feeling in feminist theory. *Feminist Theory*, *13*(2), 115-129.

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction, 34*(4), 361-388.

Philip, T. M., & Rubel, L. H. (2019). Classrooms as laboratories of democracy: The Role of new quantitative literacies for social transformation. In L. Tunstall, G. Karaali, & V. I. Piercey (Eds.), *Shifting contexts, stable core: Advancing quantitative literacy in higher education* (pp. 215-223). Mathematical Association of America Press.

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A Framework for learning about big data with mobile technologies for democratic participation: Possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning*, *18*(3), 103-120.

Rakotondravony, N. (2019, November 1). Power to the powerless: An Interview with Mona Chalabi. *Medium.* https://medium.com/nightingale/power-to-the-powerless-an-interview-with-mona-chalabi-39d73647d80a

Rick. (Host) (2021, January 11). What if Natives won? (Episode 62) [Audio podcast episode]. In *Decolonized Buffalo.* Podcast Addict. https://podcastaddict.com/episode/117657469

Roberts, J., & Lyons, L. (2020). Examining spontaneous perspective taking and fluid self-to-data relationships in informal open-ended data exploration. *Journal of the Learning Sciences*, *29*(1), 32-56.

Rubel, L. [@Laurie_Rubel]. (2020, Oct 18). "I suggest that we expand #noticewonder protocol to NOTICE FEEL WONDER ACT. the "feel" and the "act" signal our humanity, and that emotion and agency matter along with reason. could it go in some other sequence? sure. [Tweet]. Twitter. https://twitter.com/Laurie_Rubel/status/1317736725008048128

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2016a). Teaching mathematics for spatial justice: Beyond a victory narrative. *Harvard Educational Review*, *86*(4), 556-579.

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making space for place: Mapping tools and practices to teach for spatial justice. *Journal of the Learning Sciences*, *26*(4), 643-687.

Rubel, L. H., Herbel-Eisenmann, B., Peralta, L. M., Lim, V. Y, Jiang, S., & Kahn, J. B. (2022). Intersectional feminism to reenvision mathematical literacies & precarity. *Research in Mathematics Education, 24*(2), 224-248. http://doi.org/10.1080/14794802.2022.2089908

Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016b). Teaching mathematics for spatial justice: An Investigation of the lottery. *Cognition and Instruction*, *34*(1), 1-26.

Rubel, L. H., Nicol, C., & Chronaki, A. (2021a). A Critical mathematics perspective on reading data visualizations: Reimagining through reformatting, reframing, and renarrating. *Educational Studies in Mathematics*, *108*(1), 249-268.

Rubel, L. H., Peralta, L. M., Herbel-Eisenmann, B., Jiang, S., Kahn, J., & Lim, V. (2021b). Theorizing data science education: An intersectional feminist perspective on data, power, and "playing the game". Proceedings *of the 43rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA)* (pp. 217-221). https://www.pmena.org/pmenaproceedings/PMENA%2043%202021%20Proceedings.pdf

Rumack, A. M., & Huinker, D. (2019). Capturing mathematical curiosity with notice and wonder. *Mathematics Teaching in the Middle School*, *24*(7), 394-399.

Swalwell, K. M. (2013). *Educating activist allies: Social justice pedagogy with the suburban and urban elite*. Routledge.

Stornaiuolo, A. (2020). Authoring data stories in a media makerspace: Adolescents developing critical data literacies. *Journal of the Learning Sciences, 29*(1), 81-103.

Tygel, A. F., & Kirsch, R. (2016). Contributions of Paulo Freire for a critical data literacy: A Popular education approach. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3279

Van Wijnendaele, B. (2011). Social justice and the politics of emotions. *Human Geography*, *4*(2), 76-90.

Warren, C. A. (2021). *About centering possibility in Black education*. Teachers College Press.

Wilkerson, M. H., & Polman, J. L. (2020). Situating DS: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, *29*(1), 1-10.

Woodson, A. N. (2021). Afropessimism for us in education: In Fugitivity, through fuckery and with funk. In C. A. Grant, A. N. Woodson, & M. J. Dumas (Eds.), *The future is Black: Afropessimism, fugitivity, and radical hope in education* (pp. 16-21). Routledge.

Bhargava, R., Brea, A., Palacin, V., Perovich, L., & Hinson, J. (2022). Data Theatre as an Entry Point to Data Literacy. *Educational Technology & Society*, *25* (4), 93-108.

Data Theatre as an Entry Point to Data Literacy

Rahul Bhargava[1*], Amanda Brea[1], Victoria Palacin[2], Laura Perovich[1] and Jesse Hinson[1]
[1]Northeastern University, USA // [2]Univerisity of Helsinki, Finland // r.bhargava@northeastern.edu // brea.am@northeastern.edu // victoria.palacin@helsinki.fi // l.perovich@northeastern.edu // j.hinson@northeastern.edu
[*]Corresponding author

ABSTRACT: Data literacy is a growing area of focus across multiple disciplines in higher education. The dominant forms of introduction focus on computational toolchains and statistical ways of knowing. As data driven decision-making becomes more central to democratic processes, a larger group of learners must be engaged in order to ensure they have a seat at the table in civic settings. This requires a rethinking to support many paths into data literacy. In this paper we introduce "data theatre," a set of activities designed for data novices that may have limited experience or comfort with spreadsheets, math, and other quantitative operations. Through iterative co-design over three workshops, we tested and produced two activity guides for educators, building on long-standing practices in participatory theatre that center social justice and liberation. Our initial findings provide very early evidence that this approach can help these learners overcome hesitations to working with information, begin building a critical perspective when viewing data, and create emotionally impactful data stories told through theatrical performance. This prototype work suggests to us that the concept of "Data theatre" warrants further study to build a more robust understanding of its affordances and limitations.

Keywords: Data literacy, Participatory theatre, Education, Social justice

1. Introduction

Data is shaping our lives and culture in a growing number of ways, from enabling our everyday digital activities, to supporting scientific breakthroughs, to guiding decision-making at local and international scales. Yet significant work has documented the negative impacts of this trend towards datafication (Datafication is a technological trend that seeks to capture large proportions of human activity and turn them into digital and quantitative data for processing (Couldry & Mejias 2019)), such as racialized outcomes in facial-detection based policing, embedded gender bias in automated hiring, racial disparities in algorithm-based healthcare diagnosis, and more (Eubanks, 2018; Noble, 2018; O'Neil, 2016). In response to these examples, broader conversations have begun to integrate questions of justice and equity in data-centered projects (D'Ignazio & Klein, 2020; Raffaghelli, 2020; Williams, 2020).

In parallel, data literacy has grown as a major topic in educational settings across technical and socio-political domains (Prado & Marzal, 2013; Klenke et al., 2020; Ryan et al., 2019; Timmermann & Havemann, 2020). We utilize a four-part definition for "data literacy," engaging components related to acquiring data, processing and analyzing data, representing data, and storytelling with data for some purpose (Bhargava & D'Ignazio, 2015). Regions in Rwanda (UNICEF Rwanda, 2016) and Australia ("Data Literacy," n.d.) have already introduced data literacy as part of their digital literacy curriculums, and the European Union has classified data literacy as an essential skill for the future and since 2016 has implemented policies and initiatives to support its development (Carretero et al., 2017). Typical introductions to data literacy at the university level follow patterns that have long existed, privileging those with math and technology fluency and thus reinforcing existing power asymmetries in data via their structure and application (Bourdieu & Passeron, 1990). These parallel trends expose a growing need for educational data literacy practices that center principles of justice, emancipation, and participation.

Motivated by this need, in this paper we build on historical theories of education and embodiment and participatory theatre to introduce our collaboratively designed "data theatre" workshop. The first workshop activity, entitled "Embody a Dataset," invites participants to choreograph a representation of data using the other participants' bodies. The second, entitled "Make a (Data) Scene," asks teams of participants to find a story in a dataset and quickly design and perform a short scene to tell that story. We introduce motivations, summarize inspirations and related work, review the design process, document the activities, and discuss initial findings. This offers an initial case study of introducing data literacy centered around concepts of social justice and complements our larger body of work building arts-based invitations for data literacy learners.

1.1. The need for new paths into data literacy

Justice-informed pieces of building data literacy can look radically different than dominant current approaches. An important question to engage is precisely what "justice" means in our work. At a high level we embrace terms like "justice," "equity," and "capacity" to align with traditional struggles to retake power by those without. We take "social justice" to be a phrase defining a system where members of society can participate fully in their roles, achieve social mobility, and access collective support structures provided by their communities. We respect and engage the phrase's long history in Western religious thinking, and its connections to more modern movements such as Liberation Theology (Gutierrez, 1988), but do not choose to center those roots in our application.

Taking approaches built on this definition is becoming critical as more decision-making in civic settings has become data-driven. Those who do not "speak" the language of data do not have social agency in this setting; we define that as an emerging problem for democratic society. How can those without a seat at the table in civic decision-making contexts that center data be authentically invited to participate? What barriers are holding them back? This work is targeted at developing alternate paths into this set of skills to create new ways for all people to learn to "speak data" so they can engage in society critically. We are informed by a conception of three main barriers: (1) focus on statistics (2) centering of technologies (3) control over impact.

First, we find that statistical analytic processes are the primary goal of many introductions to data. Browsing a mix of educational and professional training, one finds statistics and computation are most often front and center, introduced without significant discussion of context and ethics (Oliver & McNeil, 2021). This pattern follows a long arc of abstract mathematical thinking being centered in American educational curricular guidelines (Scheaffer & Jacobbe, 2014). The current manifestation of this history is the "STEM" approach to curricular priorities, proposed as a response to forecasted economic workforce needs (National Academies, 2007). There is a history of robust critique of the form and intent of US math education (Harouni, 2015). This has led to many learners forming identities as "math" or "non-math" people; definitions only somewhat related to their actual competence (Leonard et al., 2020; Cribbs et al., 2015). As data science was introduced and claimed by the field of statistics (Gould, 2017), quantitative analytical approaches became the standard at the expense of creating a more inclusive set of pathways into the field of working with data. Any introduction to data framed through a mathematics lens immediately raises a massive barrier to the group identifying as "non-math" people.

Second, introductions to data literacy in learning settings typically rely on a techno-centric understanding of these skills, focused on quantitative data stored and analyzed via computational means. Digitization has radically altered the costs of collecting, storing, and retrieving data of any type, placing computation centrally within any data project. Related Big Data skills have been in high demand for more than a decade (Manyika et al., 2011). Focusing on software that supports algorithmic operations on large datasets privileges those with access to computational tools, which differs across social, economic, and racial divides (Jackson et al., 2008). This privileges those with power and access. In addition, Small Data continues to play a large role in social contexts (Blair et al., 2014), despite the hype behind it's "Big"-ger cousin (boyd & Crawford, 2012)

Third, datasets are too often separate from their potential applications and context of influence. Our prior work has critiqued data projects for their persistent lack of transparency and reliance on extractive collected data (D'Ignazio & Bhargava, 2015). The phrase "data is the new oil" is quite emblematic of the promise and peril of this approach (Flender, 2019). There is a relevant parallel here - just as with oil, the benefits of data projects are gleaned by those in power while the perils are experienced by those typically marginalized. This can be attributed, in part, to the fact that most data are extracted from a population or setting of study and taken to another context to be analyzed and interrogated. The analysis and impact conversations rarely integrate the subjects represented in the data, nor the communities that data-driven decisions will impact.

Prior work has begun to address this. Qualitative data can be introduced alongside quantitative to introduce non-computational analytical approaches (Henderson & Corry, 2020). Introductions can include concrete examples of alternate modes of data analysis such as collaborative meaning-making in urban planning processes (Boston Transportation Department, 2017), or story-finding and storytelling in the design of clothing and traditional fabrics (Katuli, 2019; Perovich et al., 2020). These examples show how discussion, aesthetics, design, and object construction (Willett et al., 2017) are forms of data analysis and storytelling. In addition, they demonstrate community ownership of data analysis, addressing our third main critique.

1.2. An arts-based data-theatre approach: Research questions

In this paper, we explore theatre, an arts-method, as a promising approach to address the broader need for new paths into speaking the language of data. Art-based methods have long been used to enhance learning in diverse disciplines, such as medicine (de la Croix et al., 2011), math (Hally & Sinha, 2018), and physics (Solomon et al., 2021). These methods offer learners an opportunity to tap into their kinesthetic intelligence by using their body to express or understand ideas, concepts, and experiences (Lowenfield, 1957). This may appeal to learners that embraced performance-based media for sharing information, including data-based facts (such as Tiktok and Instagram).

We describe the iterative design process, and initial findings, from a series of "data theatre" workshops we led online and in person. Our design process finds ample overlaps between the world of participatory theatre, which has a traditional focus on social justice, and our arts-based approaches to data literacy. The intersection between data literacy, justice, and the barriers enumerated above, leads us to our driving motivation: exploring how participatory theatre can help us build novel introductions to data that center on justice. Specifically, we formulated three central research questions:

- RQ1 - Literacy: In what ways can theatrical activities introduce learners to data literacy skills?
- RQ2 - Critique: How might embodying data help lead to a critical questioning of datasets and their use?
- RQ3 - Impact: Does the process of performing a data story change performer or audience perspectives on the subjects of the data?

We engage these questions in our workshop design and throughout this paper. At the same time, this early-stage work is still speculative and exploratory as we work with learners to build a practice of data theatre, developing a novel approach as a basis for further study. Future work will explore ways to more robustly assess causal relationships between workshops and data literacy outcomes, relative to more traditional data learning activities.

2. Related work

In broaching the questions of what "data theatre" might look like, what justice-oriented learning goals it might fulfill, and its potential for impact, we build on significant work in three areas: (1) education and embodiment, (2) participatory theatre practices (3) professional theatre practice.

This is guided by central design principles established in our previous work (Bhargava & D'Ignazio, 2015):

- Participation from all parties
- Learner-guided explorations
- Facilitation over teaching
- Accessibility to a diverse set of learners
- Focus on real problems in the community

In prior projects these principles have guided the development of learning experiences to design and paint community data murals (Bhargava et al., 2016) and build physical data sculptures with crafts materials (Bhargava & D'Ignazio, 2017). This new foray into activities for data theatre introduces our work to a rich history of education through embodiment, and introduces a new performative output for the learning activities we create.

2.1. Education and embodiment

Research in education and embodiment surfaces the importance of emancipation through critical data literacies and points to the ways in which physical movement and action may alter understanding, perception, and memory.

Many pedagogical approaches have embraced justice and equity at their core. We take strong inspiration from Dewey's foundational connection between education and democracy (Dewey, 1903) and Freire's approach to emancipatory education (Freire, 1968). In the context of data literacy, we specifically posit that the application of this idea to modern day involves both understanding the oppression of data processes integrated into civic governance and taking back power to use data in service of collective justice (Couldry & Mejias 2019). Functionally, this builds on research into how learning can be more effective when the subject is interrogating issues that matter in their communities (Zeidler et al., 2005). From this work we draw an approach that works with local community data, provides creative freedom to find narratives in it, and focuses on potential real-world impact of the process.

An emerging definition of "critical data literacy" applies Freire's approach to the concept of data literacy (Tygel & Kirsch, 2015; Sander, 2020). This conceptualization offers links to previous movements in critical media literacy and digital literacy (Knaus, 2020). It also provides a theoretical framework to underpin the design of activities and interventions for educational contexts that seek to raise awareness of the datafied society around us and put data in the hands of the communities it is too often extracted from. Critical data literacy embeds a hope for empowerment through activism and social change (Gutiérrez, 2018). The approach of critical data literacy offers us theoretical framing within which to situate our activities, and a set of related projects that our work connects to.

Finally, we find inspiration in embodied approaches to learning. These build on understandings of learning as situated in social context, activity, norms, experience, and physicality (Lakoff & Johnson, 1999). Indeed, scholars have acknowledged the power of embodied experiences to shape learning (Höök, 2018) and promote empathy (Yap, 2016). Going further, we embrace the idea that all forms of knowing are rooted in the "somatic phenomenon" of the learner - the physical experiences they undergo and create while in learning settings (Abrahamson & Lindgren, 2014; Malinverni & Pares, 2014). Data science educators have built on this type of definition to offer activities such as bodily data sorting and physical infographics (Sommer & Polman, 2017). The latter two are examples of participatory simulations - opportunities to socially construct an understanding of phenomenon by embodying it. In this practice the act of "knowing" itself is participatory and shapes outcomes (Barab et al., 1999). The emerging concept of "data feminism" embraces this, suggesting data embodiment and data "visceralization" be valued as ways of knowing data (D'Ignazio & Klein, 2020).

Furthermore, our definition of justice comes together with a motivation in an attempt to embrace "epistemological pluralism" in the domain of data science. As defined by Turkle and Papert, epistemological pluralism is an "acceptance of multiple ways of knowing and thinking," with each "equally valid on its own terms" (Turkle & Papert, 2015). Their conceptualization connects to gender rights in learning settings, respect for previously dismissed concretized learning, and notions of body syntonicity. We are inspired and motivated to create a data science that engages and values epistemological pluralism, moving beyond a belief that algorithmic and formal thinking is naturally superior. Embodied approaches to building data literacy are one piece of this puzzle.

2.2. Participatory theatre

Our work draws on the long-standing focus on liberation in participatory theatre and the many practical activities built by expert practitioners in this space over decades of contextual work with communities.

The approach of "epic theatre" is a common root for many of these practices, built around the idea that theatrical practices can provoke the audience to reflect on the world as it is, rather than suspending disbelief and entering an imagined world via the performance (Styan, 1981). Theatre of the Oppressed, a form of theatre designed for social and political activism, explores oppression in all forms - internal / external, individual / societal, and more (Boal, 1993). Verbatim theatre (Catchesides, 2020), and image theatre (Farmer, 2014) are all examples of other approaches that have grown from this common theoretical core and been successfully used for empowerment in communities. Playback theatre, where participants share experiences which are then "played back" to the broader group through, has been used to promote compassion and understanding in medical education (Salas et al., 2013), healing and reconciliation for post-war communities (Dirnstorfer & Saud, 2020), mental-health recovery (Moran & Alon, 2011), and healing for adolescents in refugee camps (Edelbi, 2020). It aims to create a deeper, multi-dimensional understanding of experiences — building empowerment and engagement in decision-making.

These approaches interrupt the theatrical experience and charge participants with decision making, analysis, and sharing judgment, blurring the lines between actors and spectators. The audience may have strong visceral and emotional reactions to a piece, but they are more likely to take social action when they are invited to act in the experience of the theatre. Functionally, these approaches work in support of critical thinking and participatory engagement via what Brecht refers to as a "dialectical" approach, focused on meaning making through engagement of potentially opposing perspectives (Mumford, 2008).

These forms of theatre align with our goal to create participatory spaces that bring people together to learn how to speak and embody data, and leverage this knowledge towards goals within their communities. The element of shared discovery and storytelling, structurally influenced by the perspectives and experiences of participants, can create a sense of control and responsibility amongst the collaborators (PARCOS, 2020). In the words of social science dramatists and researchers Ross Gray and Christina Sindig, “A drama that emerges from such a process is very much a negotiated settlement, a collective achievement” (Gray & Sinding, 2002). Participant audiences reconcile with their experiences through story building, a process that connects strongly to the definition of critical data literacy.

2.3. Professional performances

Finally, we looked to professional theatre practice for inspirations of performative data productions in action. The 2015 piece *A Sort of Joy (Thousands of Exhausted Things)*, created by the theatre troupe Elevator Repair Service and the data visualization group Office of Creative Research (Thorp, 2015) uses metadata from the New York City Museum of Modern Art’s collection to create a performance that highlights the gender of artists whose pieces are in the collection. Performed at the MoMA itself, the piece centers around faux visitors reading out names in cadence. A circle of men stands with three women apart. The men speak other male names for several minutes, until finally the women speak for the first time - saying “Mary” to highlight the gender gap in museum acquisitions. The entire structure of the piece pulls the audience into the act of questioning the gender diversity of the MoMA’s collection though experiencing the gaps in the representation via their eyes, ears, and body as it awaits the next time the women speak.

To further explore and understand professional practice, we conducted two interviews with professionals in the field of theatre. Our first interview was with Christopher Ellinger, Founder and Director of True Story Theater, a nonprofit playback theatre company that offers improvisational performances and workshops to community groups, businesses, and individuals. Ellinger shared a wealth of experiences around how to help participants develop self-awareness and promote empowerment, reconciliation, and decision-making. Our second interview was with Frederica Fragapane, an award-winning Italian information designer, who researches the mutualistic relationship between data visualization and performance. Some of her work has integrated traditional practices with projected data visualizations that reflect what is happening in the play - such as network diagrams of character interactions projected behind an unfolding scene (Fragapane, 2017). Her work treats theatrical production as an artifact for interpretation and representation via standard data visualization approaches; an alternate conception of what data theatre could be.

3. Workshop and activity design

With this theoretical grounding, and related work across disciplines, we used an iterative process to develop a set of data theatre activities centered on justice. Over three workshops we iterated on the design of the activities with students and qualitatively assessed the experiences against our research questions and motivations. Each workshop focused on one of our driving research questions, while touching on all of them. Our workshops are also informed by interviews with two professionals in the field of participatory theatre. In this section we summarize our methods, workshops, and observations.

3.1. Workshop structure and development

The three workshops included (1) an initial exploratory workshop focused on RQ1- Literacy, (2) a prototype theatre-focused workshop focused on RQ2 - Critique, and (3) a single session in-class workshop focused on RQ3 - Impact (Table 1).

Table 1. Workshop format and details

Workshop #1	Workshop #2	Workshop #3
Exploring Movement	Prototype Data Theatre	In-Class Data Theatre
May 2019	December 2020	February 2021
In-person	Virtual (via Zoom)	In-person
Graduate students	Theatre undergraduates	Theatre undergraduates
15 participants	10 participants	12 participants
Self-selected	Self-selected	Class members
Focused on RQ1 - Literacy	Focused on RQ1 - Critique	Focused on RQ3 - Impact

Each workshop was built around an introductory experience to set tone and activate participants’ bodies, followed by two in-depth activities. Workshops were 1.5 hours, with interspersed reflection discussions between activities. Discussions with participants during the sessions were facilitated as semi-structured focus groups, consisting of thematic prompts from the facilitator followed by open time for reaction and engagement.

At each workshop we utilized the same data handouts. These included

- A 1-page handout with two charts about ice cream consumption in the US: a line chart of per-capita consumption by year, and results of a small survey about favorite flavors.
- A 4-page handout about issues of food security in a local community, based on surveys administered by a non-profit. This included items such as tables summarizing time spent getting groceries, quotes about barriers to acquiring appropriate food, and bar charts of population demographics.
- A 4-page handout about healthy eating habits from a government-run local early childhood support program. This included tables of population demographics, bar charts indicating family size and country of origin, quotes about preferred patterns of eating, and more.

Between sessions we iterated on the activities and incorporated virtual or physical constraints in addition to participant feedback. In Workshops #1 and #2 we attempted to engage participants as co-designers of the experience; commenting on, and suggesting ideas for, how to run activities in this domain. This deliberately reinforced our approach to justice and empowerment, not just in terms of product but also the processes we undertook. Due to the speculative nature of this work, and our desire to have participant feedback shape next steps, we hoped this format would allow space for topics and reflections we did not anticipate to emerge.

3.2. Data collection and analysis

Multiple streams of data were collected at each workshop. Each workshop was video recorded and automatically transcribed to support later evaluation work. During the workshops, members of the research team took on one of three roles:

- Facilitator: the leader responsible for introducing activities, monitoring progress during breakout sessions, and prompting reflection.
- Non-participant observer: a “fly on the wall” tasked with taking notes about the interactions and reactions of participants during the session itself, including notable participant statements and reflections.
- Participant: worked alongside students and contributed comments on the social dynamics of the session after the fact.

We held a short debrief after each workshop to share comments and reflections, with collective note-taking happening in a shared online document.

We conducted evaluation through qualitative review of these outputs. One team member reviewed the video transcription, observer notes, and debrief notes to synthesize high level findings. This involved looking for recurring observations based on connections to our three research questions and extracting key quotes from participants that were marked or mentioned by researchers in any of the notes taken. We took a hybrid approach to doing this qualitative descriptive coding, based on both our primary research questions and topics that emerged within the transcriptions and notes themselves. We deemed this appropriate because while we do have some key research questions, this work was also highly exploratory in nature. Findings from this qualitative data analysis were brought to the larger research team for reflection and further discussion.

3.3. Workshop #1: Exploring movement and data literacy

Our first workshop was an exploration of movement and data designed to surface initial reactions and reflections (Table 1). The driving research question for this workshop was RQ1 - Literacy. We structured it to evaluate what kind of basic data literacy skills could be explored via movement and embodiment.

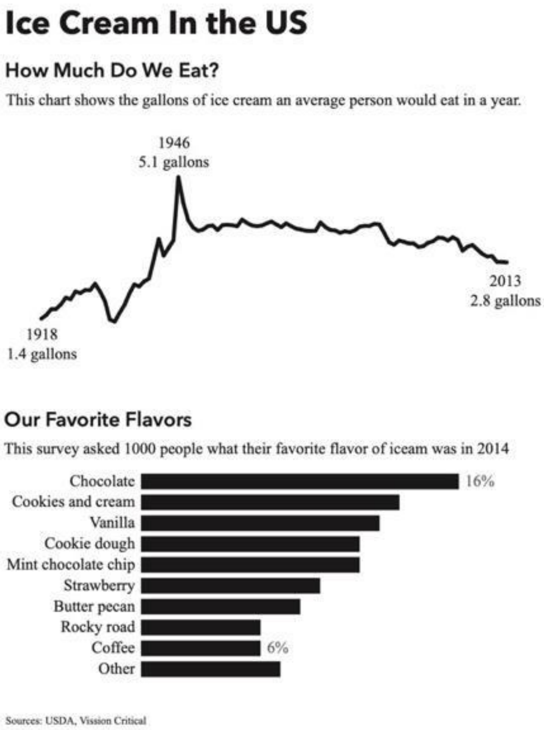
3.3.1. Activities

The two central activities were (1) a “puppeting” experience and (2) a collaborative design of a short performance in small teams.

Activity #1 – Puppeting: This activity was inspired by image theatre (see Section 2.2), where participants “sculpt” bodies to embody perspectives, and augmented by movement cards (Figure 1) we developed through a preliminary review of movement research. In this activity, each group of four received a movement card (Figure 1) and a dataset (Figure 2). Each pair was invited to review the data handout, about ice cream consumption in the US (Figure 3), and then to instruct the other pair on positioning their bodies to express the data, constraining their actions to the movement prompt card they received. Then the roles switched.

Figure 1. A sample of the movement cards provided to participants

Figure 2. The ice-cream data handout



Activity #2 – Collaborative design of a short performance: Each group received one of two multi-page qualitative and quantitative data handouts; one focused on local food security data, the other on a local program around healthy eating. The groups were instructed to review and discuss the data handout and develop a series of movements to represent a story that they saw within it. After about 15 minutes of work time, the groups were re-engaged to share their performances (if they wished to).

Figure 3. Participants reviewing data and planning their performance

3.3.2. Observations and reflections

Some of the pieces created included:

- Puppeting bodies to mimic the curve of a line chart showing ice cream consumption over time.
- Puppeting peers to mime ice cream being scooped and served based on select rates of consumption.
- A performance representing hunger via painful expressions and doubling over, and highlighting barriers to food access by performing the wait at a bus stop.

The overall workshop was both playful and reflective. We were successfully able to create an atmosphere of just trying things out; as seen in comments about “we were having fun,” “that was really fun,” and “all that data stuff was a lot of fun.” The last one is particularly telling for us, because it suggests that the participant did not expect working with data to be fun.

Specifically related to RQ1 - Literacy, we saw initial evidence that participants were engaging with some core areas of data literacy: encodings, representation, and editorial choices in storytelling.

The two puppeting examples shared above indicate that participants were exploring a variety of modes for representing data in physical form. The first group commented that they were recreating the chart “through bodily movements like a one-to-one translation.” This suggests to us that this team was exploring how to bring the existing 2D representation into a 3D space with their bodies. Perhaps their meaning-making through mimicking the chart might be like learning to read a graph? On the other hand, the second example, miming ice cream, to us was akin to using symbols to represent data, as in pictographs. They were using the literal act captured in the data, eating ice cream, to represent to the audience a story they saw in the data. This reflects evidence from other work where we found early novice data literacy learners often draw symbolic depictions to represent data (Bhargava et al., 2021). Taken together, these two examples show participants engaged in explorations of body-based representation that brings them into contact with questions asked in more traditional data literacy introductions. They stated how the intention to “to use the movements to explain the data,” and “used different movements for different quantities.” The visual encoding space of traditional charts is well-document and explored, but the encoding and representations possible via embodiment merit further investigation.

We were also provoked into questions about predisposed notions of “analysis” and “embodiment” being separate tasks. One participant noted “it’s really weird to think about how to turn numbers into something... having no scales changes things.” This offers a potentially interesting hypothesis that relates to prior work reviewed on education and embodiment (Lakoff & Johnson, 1999; Höök, 2018; D’Ignazio & Klein, 2020), specifically that non-graph-like embodied ways of understanding were being dismissed by participants even when prompted to create embodied representations of knowledge. Interestingly, one participant went further and pointed out that the limits of embodiment also mattered - “what are we trying to do… what can we not do physically?” Relatedly, another commented that it “was really interesting to think about how you make scales different based on the type of visualization”; perhaps demonstrating an ability to reflect on chart design decisions after the physical act of embodying the data.

The performance activity showed participants wrestling with questions of turning a data into a story. One noted that “it is really hard to represent all of the data,” a reflection related to questions of what data is included in a story. When prompted, one participant more directly commented that “I found the hardest part was to create a bit of a narrative there… how do we take the different parts of the data and make a story out of it.” One participant in response commented that “even with the same graph, we interpret” - perhaps pointing to an early understanding that any data can contain a multitude of stories. Another pointed out that as an audience member you “don’t have a context for the data… you’re almost looking more closely, trying to process every movement.” Storytelling with data is a difficult skill, but these comments suggest that the performance activity in particular was leading participants to wrestle with it. This provides early evidence that our two activities were helping participants explore some fundamental pieces of data literacy - specifically how to represent data and turn it into a story.

On a final note, we found most people focused on acting out stories from the data as skits, while a few translated the data into more abstract dance-like movement. Their skits were very narrative in their structure, while the dance seemed to be evoking emotions they saw in the data. This drove a shift in our project structure, offering two paths: (1) diving into data and dance (publication forthcoming), and (2) exploring data and theatre (described in this paper).

3.4. Workshop #2: Prototype data theatre

We specifically designed the next workshop to use practices of participatory, socially engaged theatre to build data literacy in a group of participants that did not see themselves as “speaking data.” We built on a Brechtian approach to theatrical practice and invited participants in as co-designers of the activities; after each activity we asked them to take off their “participant hat” and comment on the design of the activity itself with their “facilitator/educator hat.” Building on what we observed at the first exploratory workshop, and our learnings from the field of participatory theatre, our driving research question for this workshop was RQ2 - Critique.

3.4.1. Activities

The two central activities were (1) a collective “puppeting” experience and (2) creating short data skits in small groups.

Activity #1 – Collective Puppeting: Here we extended the puppeting activity to the group at large - pairs of participants designed ways to showcase something they saw in the ice-cream data by instructing the rest of us about what to do with our bodies. This connected to our goal of breaking down the divide between audience and performers - inviting the entire audience to make meaning by performing a piece. The reflection discussion afterward focused on questions more targeted at critical thinking.

Activity #2 – Data Skits: After discussing what makes up a “story” (versus just raw data) we introduced the idea of finding a story and designing a 2–3-minute skit that tells that story. To dig into RQ3 - Impact a bit more we asked them to focus on conveying one key emotion to their audience. They were given about 15 minutes for this activity and invited to perform their piece after that.

3.4.2. Observations and reflections

Some of the pieces created included:
- Puppeting the group to compose a line chart with their arms across a grid of video boxes in Zoom.
- Puppeting one participant at a time to represent the popularity of a flavor of ice cream by trying to represent it via movement and sound.
- A skit comparing time spent getting to a grocery store by different populations via pretending to drive a car.
- A skit acting out a debate between two fictional community members about whether the data about cultural eating norms was reductive, stereotypical and offensive.

Overall, we were pleasantly surprised at the engagement from participants despite the constraints of the virtual setting. We hypothesize that leading silly introductory activities within the boxes on Zoom created a sense of failure and awkwardness being a normal part of the process, freeing participants to be slightly more uninhibited in their ideas. While reflecting on this first activity, some participants shared that they were a little lost at first, another commented that "once we got into it, it was really really fun." Again, this reinforced a finding from Workshop #1, specifically that embodying data via our activities was a fun process.

The car driving skit provided an evocative example of using theatrics to create space for questioning in the audience (similar to the *A Sort of Joy* piece described in 2.3). The awkward silence of the final driver continuing to drive for far too long created space to question why their drive was so much longer.

A number of comments offered early evidence that we are creating a new pathway for learners who perceive themselves as "non-numbers" people. An incidental conversation as the workshop was starting led to many participants self-identifying as "non-numbers people." This binary identity formation in relation to math creates a real barrier to entering domains that learners assess as numbers-heavy (Leonard et al., 2020). In future workshops we would more robustly capture this in a pre-survey, providing an informal qualitative indication of how the group felt. In addition, we could embrace Leonard's notion of "naming discomfort" as a potential intentional response to try break down any perceived barriers felt by participants.

The debate skit mentioned above provided our strongest example of participants engaging in critical questioning of datasets and their use (as described in RQ2 – Critique). They challenged the assumptions they saw in the data. Based on our decade of experience running data literacy activities we found this quote notable - critiquing the data like this rarely happens with novices. The fact that this group did just that suggests that we were able to create an environment where they felt comfortable engaging their critical thinking skills despite the power dynamics at play within the workshop. In addition, some of this early evidence related to our commentary on control of impact (section 1.1). Our choice to use local data resonated with a number of the participants. One noted that our approach "allows communities to redefine what is a hard-tangible fact with emotions" and lets "the audience have freedom to take from it what they are going to take from it." These examples suggest that our Freire-inspired approach to designing with community data with impact might have successfully introduced the basics of the critical data literacy concept introduced previously.

At the process level, one helpful comment suggested that calling them data "skits" was inappropriate; the term implies humor, and this was not engaging in humorous approaches.

3.5. Workshop #3: In-class data theatre

Our third workshop was a single session in one of the author's in-person Movement course for undergraduate theatre majors. As with workshop #2, our main goals revolved around testing the activities to determine how they support helping a new group of participants start to move from data to story through the design of participatory performances.

3.5.1. Activities

The two central activities were (1) a prompt to "embody a dataset" and (2) a prompt to "make a (data) scene."

Activity #1 – Embody a Dataset: We did not substantially alter this activity.

Activity #2 – Make a (Data) Scene: Building on our learning from Ellinger of True Story Theatre (see 2.3), we added a bit more intentionality in how we introduced the data. Using the same handouts, we took more time to address the fact that they might connect to negative experiences of people in the room and we attempted to create space to honor and digest that. Additionally, we began using the term "scene" in place of "skit."

3.5.2. Observations and reflections

Some of the pieces created included:

- Embodying data about ice cream consumption over time by directing the class to move around the space and hum, while varying volume and speed based on the data instructions.
- Embodying data about consumption levels via standing at different heights.

- A scene making physical all the obstacles that a disadvantaged person faces in accessing food for her family, from time, to distance, to money, in a journey to feed her children.
- A dramatic scene where a character was taunted with healthy food and other resources that were too high above her head or too fast for her to catch.

The “embody” activity offered two interesting notes. The first related to sound and movement - this was the first intentional use of encoding via sound that we saw. We found it encouraging that these activities led participants to a mode of data representation that has only recently gained in popularity - sonification (Lenzi & Ciuccarelli, 2020). The second relates to “binning” - the process in statistics of grouping a set of discrete data into larger categories. In this specific case, the team quickly averaged ice cream consumption levels by decade to instruct participants how to stand. In more traditional data literacy programs, we often see learners introduced to the idea of binning and how it can alter the results of an analysis (based on bin size). Here we saw the participants come to the idea themselves and used the opportunity to discuss the idea and concerns during the reflection after they presented it. These both suggest interesting learnings related to the analysis and representation pieces of data literacy.

Regarding RQ3 - Impact, we saw early indications of participants reflecting on the subjects of the data in meaningful ways. One group noted that the “make a scene” activity helped them get around the dehumanization that can occur in traditional representations, saying specifically that “numbers / statistics are not just numbers, they represent people going through this.” A viewer of that scene said it “pulled me in because it matched something that happened to me in my life.” These comments help us dig into some of the driving questions related to how performing a dataset could change perspectives on those represented by the data, and potentially build empathy for their situation. Furthermore, one participant started to get at a potentially causal link that merits more investigation, stating that their “emotional response to the data was a lot stronger when we embodied the data.” However, another commented that one weakness is how “the paper visual representation is stronger because you can reflect back and look at it … it is locked in time.” These suggest the participants were engaged in the act of embodying the data and were reflecting on its affordances; there may be aspects such as recall that are not supported through embodied representation.

3.6. Synthesizing two activities

Based on our interaction and reflections from the 3 separate workshops we synthesized our learnings into two data theater learning activities for novice audiences. We offer these as the foundation of an approach to data theatre in service of equity and engagement in learning settings that focus on justice. The “Embody a Dataset” and “Make a (Data) Scene” activities are documented in activity guides for educators attached as supplemental material to this paper.

4. Synthesis and discussion

In this section we highlight findings and reflections across workshops and research questions. In particular, we synthesize and discuss our initial findings in relation to each of our guiding research questions. With our focus on undergraduate theatre students as primary participants, findings may not be generalizable to other audiences we plan to work with, such as community organizations. However, the findings do suggest an initial foundation on which to build and are applicable for others wishing to work with this same or related audiences. Our early observations suggest that data theatre is a novel entry point into data literacy, worthy of more study (as detailed in 5 below).

4.1. Creating many paths for many learners (RQ1 - Literacy)

Our initial workshops and prototype activities suggest that data theatre can introduce participants to several core data literacy skills - reading data, picking representations, and creating stories.

Our activities were appealing to audiences that might not be drawn to spreadsheets, math, and the computational analytic tools typically used in introductory data literacy sessions. Participants associated the term “data” with other words such as “information,” “numbers,” “research,” “charts,” and “technology.” When prompted to represent “data” with a sound of movement, we heard computational bleeps, bloops, and mechanized motions. In addition, we found in Workshops #2 and #3 that our participants did not identify as “math people” (Leonard et al., 2020). One participant noted, "You are translating your math mind into a creative mind, into a story mind, into a people mind." Another participant echoed that "(we were) looking at a piece of data and the numbers and quantitative stuff, but what's the story here." A third said that "I think that is part of finding the stories, you have to translate." These comments suggest that participants were building basic data literacy skills, specifically in the realms of representation and storytelling. Our responses could employ Cribbs' proposed pathway from "performance" to "recognition" to "mathematics identify" more intentionally in response to these types of comments about engaging different parts of one's brain (Cribbs et al., 2015). Combined, these observations deepen our motivation to create pathways into data literacy for learners who are not attracted to the existing computation-centered introductions, specifically related to our concern about citizens' abilities to join in civic discussions that are increasingly data-centered.

This returns to our goal of validating multiple ways of knowing in the field of data science. Specifically informed by the concept of epistemological pluralism (Turkle & Papert, 2015) and data feminism (D'Ignazio & Klein, 2020), data theatre can serve as an entry point for learners that take a different approach to thinking through data. For instance, while many introductions to data begin with aggregating a spreadsheet, we saw many of our participants begin by trying to understand one data point. Perhaps like the story of Anne that Turkle and Papert's share, where the "bricoleur programmer" covers her bird like a painter might, data science needs to embrace solutions that are informed by approaches from other domains, even if they seem "sub-optimal" or "inefficient" to experts. Our "data point explorer" should also be welcomed as having a valid way of exploring a dataset. Another connection here is to the idea of kinesthetic memory as another way of knowing (Lowenfield, 1957). Participants in our workshops were using their body to process, understand, and express data concepts and interpretations.

4.2. Building critical data literacy (RQ2 - Critique)

In this context, we see data theatre as a promising approach that could help engage a critical frame of mind about the data itself, though this work did not engage critique of data's use directly.

It is illuminating to bring this discussion back to Tygel and Kirsch's (2015) conceptualization of critical data literacy. Our participants pointed out that "statistics and data sometimes generalize experiences and are not telling the whole story just by looking at the numbers," and "using reflection and performance shows the story underneath it that the data attempts to bury." These comments suggest that, as Tygel and Kirsch (2015) define it, our participants are seeing data as an "output of a social process." In our second workshop, one participant noted that "I was really self-conscious about being critical of the data set," but then they proceeded to despite that hesitation (creating the data debate described in section 3.4.2). In reference to Tygel and Kirsch's (2015) definition again, we argue this group demonstrated data "manipulated based on explicit objectives" and including a "social evaluation of what message should be transmitted." Specifically, we found participants were willing to interrogate and question data when asked to physically embody it. As one participant noted in the Workshop #3, you "bring your own interpretation or experience to the performance." A data feminist might suggest you do the same with data - "consider context" is the sixth principle of data feminism (D'Ignazio & Klein, 2020).

We saw less evidence that we effectively prompted critical reflection about the use of data in social settings as disempowering or emancipatory. This has become more important to engage as large companies with massive technical capacity continue to use data-based systems to impact far larger swaths of the population, with more significant impacts on the historically marginalized (Noble, 2018). If we create new paths to engage populations in civic data-driven decision making, our definition of justice indicates it must be in service of those with the most potential to be harmed by the decision. Our prompts were not designed to help participants consider the impacts of data use in a community.

4.3. Creating impactful embodied experiences (RQ3 - Impact)

These initial workshops lead us to believe that our "Make a (Data) Scene" activity can rehumanize data and potentially build empathy and engagement for participants, but we are left with more questions about impacts for the audience.

One consistent reflection in our work was that the theatrical embodiments of these data points help participants rehumanize the data. Typical introductions to data are decontextualized from the subjects of the data, and their points of impact (Oliver & McNeil, 2021), a critique we have made in other work about introductions to data literacy (D'Ignazio & Bhargava, 2015). We argue that with a justice-focused lens it is centrally important to integrate the lived experience of those represented within a dataset, and the context of its potential implications. Participants reflected that the scene activity "grounded it in the actual people" and "got me thinking about the people behind the data, behind the numbers." They were "telling the story of the people behind the data." This provides early evidence that embodying data may have helped these learners remember that data often represents people and experiences. This work allows them to consider their own experiences and knowledge in relation to the communities the data is about — key steps for building empathy.

Furthermore, we heard some reflections suggesting that our activities created impactful reflective experiences for the participants themselves. For instance, responding to a question about the datasets, one participant noted that it "pulled me in because it matched something that happened to me in my life." This comment reinforces findings from the field of teaching "data for good," where educators have found that learners get drawn in by working on datasets about problems they have experienced or are interested in (Bhargava, in press). This also relates to comments from our interview with Ellinger of True Story Theatre, about creating space to process seeing oneself in the data — engaging the audience in empathetic reasoning.

5. Conclusions and next steps

In this paper we discuss the motivations, inspirations, process, and findings from the creation of a set of "data theatre" activities in a higher education setting. We offer the activities as an alternate entryway to building a critical data literacy, one that builds on processes rooted in questions of justice and equity, decentering technology and inviting sets of learners who are not engaged through current approaches. We find early evidence that the data theatre approach is effective at helping participants build some data literacy skills without introductions to statistics and computational tools, reflect critically on datasets and their intended use, and engage in emotionally meaningful embodied performance of data stories.

In a datafied world, we argue that democratic societies are morally responsible to govern through understandable mechanisms. Data theatre offers one path to making data more understandable, particularly for learners without access to technology, or who do not take to spreadsheets. We hope the activity guides included as appendices contribute concretely to other educators' toolboxes. Theatre based introductions to data literacy can be an effective entry point into data literacy that complements other approaches and potentially lowers barriers to certain content based on their embodied nature.

5.1. Next steps

In an educational setting and a community setting, these introductions should serve as tools of empowerment for populations typically left out of data-centered decision processes. We hope to move beyond the college classroom in future work, trying these activities out with community groups that work on issues of justice that involve data. The coronavirus pandemic limited our ability to start with these groups, and in-person workshop with them would remove the barriers of access to bandwidth and computers for online Zoom sessions.

These initial observations and examples present us with numerous potential next steps focused on exploration, formalization, and robustness. For instance, we could administer pre- and post-surveys at workshops asking participants about their confidence levels vis-a-vis working with data. This would ascertain whether a single experience with these activities creates a short-term impact on self-assessed abilities, and perhaps build an evidence base for overcoming skills-based identity barriers to working with data. Similarly, to explore the challenge of building empathy with subjects of data, we could more directly ask learners how connected they feel to the data and the performance they created.

Particularly related to creating impact on the audience, we believe there is more theoretical grounding from theatrical practice to explore. One path forward would connect more directly to the concept of "kinesthetic empathy" - the idea that audiences experience empathy by watching a performance (Martin, 1975; Reynolds & Reason, 2012). Another relates to further exploration of feminist theatre practice and the idea of creating spectacle via performance in public place. The work of the Women's Street Theatre Group at the 1970 Miss World competition is an evocative example (Cowley et al., 1971). In addition, we could engage more directly with the Brechtian "dialectical theatre" approach of interrupting performance to engage the audience. A participant noted we could "think about not only how it can abstract from a situation but also how you can re-invite into that situation."

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In *The Cambridge handbook of the learning sciences* (2nd ed., pp. 358-376). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.022

Barab, S. A., Cherkes-Julkowski, M., Swenson, R., Garrett, S., Shaw, R. E., & Young, M. (1999). Principles of self-organization: Learning as participation in Autocatakinetic systems. *Journal of the Learning Sciences, 8*(3-4), 349-390. https://doi.org/10.1080/10508406.1999.9672074

Bhargava, R. (in press). Teaching data that matters: History and practice. In J. E. Raffaghelli (Ed.), *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead*. Springer.

Bhargava, R., & D'Ignazio, C. (2015, April). *Designing tools and activities for data literacy learners* [Paper presentation]. Data Literacy Workshop at WebScience 2015, Oxford, UK. https://voragine.net/img/2021/02/DataLiteracyRahulCatherine.pdf

Bhargava, R., & D'Ignazio, C. (2017, June). *Data sculptures as a playful and low-tech introduction to working with data* [Paper presentation]. Designing Interactive Systems, Edinburgh, Scotland. https://hdl.handle.net/1721.1/123453

Bhargava, R., Kadouaki, R., Bhargava, E., Castro, G., & D'Ignazio, C. (2016). Data murals: Using the arts to build data literacy. *The Journal of Community Informatics, 12*(3), 197-216. https://doi.org/10.15353/joci.v12i3.3285

Bhargava, R., Williams, D., & D'Ignazio, C. (2021). How learners sketch data stories. In *2021 IEEE Visualization Conference (VIS)* (pp. 196-200). IEEE. https://doi.org/10.1109/VIS49827.2021.9623299

Blair, D., D'Ignazio, C., & Warren, J. (2014). *Less is more: The Role of small data in 21st century governance*. Public Lab. https://publiclab.org/n/11421

Boston Transportation Department. (2017). *Go Boston 2030: Vision and action plan*.

Boal, A. (1993). *Theatre of the oppressed* (C. A. McBride, Trans.). Theatre Communications Group.

Bourdieu, P., & Passeron, J.-C. (1990). *Reproduction in education, society and culture* (R. Nice, Trans.; 2nd ed.). Sage Publications.

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Carretero, G. S., Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital competence framework for citizens* (EUR 28558 EN). Publications Office of the European Union. http://svwo.be/sites/default/files/DigComp%202.1.pdf

Catchesides, K. (2020). *Verbatim theatre*. Stockroom. https://www.stockroom.co.uk/old-pages/verbatim-theatre/

Couldry, N., & Mejias, U. (2019). *The Costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Cowley, S., Filner, C., Golditch, R., Hope, E., Howard, P., Hoyland, W., Lamb, K., Linebaugh, B., Melamed, A., Reid, T., Layla, & Simpson, A. (1971, March). News and notes. *Shrew: Women's liberation workshop, 3*(2).

Cribbs, J. D., Hazari, Z., Sonnert, G., & Sadler, P. M. (2015). Establishing an explanatory model for mathematics identity. *Child Development, 86*(4), 1048-1062. https://doi.org/10.1111/cdev.12363

Data Literacy. (n.d.) *Queensland curriculum and assessment authority*. https://www.qcaa.qld.edu.au/p-10/aciq/frequently-used-resources/data-literacy

de la Croix, A., Rose, C., Wildig, E., & Willson, S. (2011). Arts-based learning in medical education: The Students' perspective. *Medical Education, 45*(11), 1090-1100. https://doi.org/10.1111/j.1365-2923.2011.04060.x

Dewey, J. (1903). Democracy in education. *The Elementary School Teacher, 4*(4), 193-204.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to building big data literacy. *Bloomberg data for Good Exchange 2015*. http://www.kanarinka.com/wp-content/uploads/2021/01/DIgnazio-and-Bhargava-Approaches-to-Building-Big-Data-Literacy.pdf

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. The MIT Press.

Dirnstorfer, A., & Saud, N. B. (2020). A Stage for the unknown? Reconciling postwar communities through theatre-facilitated dialogue. *International Journal of Transitional Justice, 14*(1), 122-141. https://doi.org/10.1093/ijtj/ijz038

Edelbi, K. (2020). *Using Playback theater with adolescents in refugee camps in Palestine to tell their stories* (Publication No. 27833283) [Doctoral dissertation, Lesley University]. ProQuest Dissertations and Theses Global.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Farmer, D. (2014). *Image theatre*. Drama Resource. https://dramaresource.com/image-theatre/

Flender, S. (2019). *Data is not the new oil*. Towards Data Science. https://towardsdatascience.com/data-is-not-the-new-oil-bdb31f61bc2d

Fragapane, F. (2017, April 26). *Data visualization and theatre: A Story of mutualism*. Medium. https://medium.com/@frcfr/data-visualization-and-theatre-a-story-of-mutualism-5e199009cf4b

Freire, P. (1968). *Pedagogy of the oppressed*. Routledge.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25.

Gray, R. E., & Sinding, C. (2002). *Standing ovation: Performing social science research about cancer*. AltaMira Press.

Gutierrez, G. (1988). *A Theology of liberation: History, politics, and salvation* (C. Inda & J. Eagleson, Trans.; Revised ed.). Orbis Books.

Gutiérrez, M. (2018). *Data activism and social change*. Palgrave Pivot. https://doi.org/10.1007/978-3-319-78319-2

Hally, T., & Sinha, K. (2018). SHINE for girls: Innovating STEM curriculum with dance. *Childhood Education*, *94*(2), 43-46. https://doi.org/10.1080/00094056.2018.1451689

Harouni, H. (2015). Toward a political economy of mathematics education. *Harvard Educational Review*, 85, 50-74. https://doi.org/10.17763/haer.85.1.2q580625188983p6

Henderson, J., & Corry, M. (2020). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*, *14*(2), 232-244. https://doi.org/10.1108/JRIT-11-2019-0074

Höök, K. (2018). *Designing with the body: Somaesthetic interaction design*. MIT Press.

Jackson, L. A., Zhao, Y., Kolenic, A., Fitzgerald, H. E., Harold, R., & Von Eye, A. (2008). Race, gender, and information technology use: The New digital divide. *CyberPsychology & Behavior, 11*(4), 437-442. https://doi.org/10.1089/cpb.2007.0157

Katuli, M. (2019, January 2). Young artists use fashion and data to promote dialog on sexual health. *Data Zetu.* https://medium.com/data-zetu/young-artists-use-fashion-and-data-to-promote-dialog-on-sexual-health-517429662ec2

Klenke, C. M., Schultz, T. A., Tokarz, R. E., & Azadbakht, E. S. (2020). Curriculum data deep dive: Identifying data literacies in the disciplines. *Journal of EScience Librarianship, 9*(1), e1169. https://doi.org/10.7191/jeslib.2020.1169

Knaus, T. (2020). Technology criticism and data literacy: The Case for an augmented understanding of media literacy. *Journal of Media Literacy Education, 12*(3), 6-16. https://doi.org/10.23860/JMLE-2020-12-3-2

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The Embodied mind and its challenge to western thought*. Basic Books.

Leonard, A. E., Bannister, N. A., & D'Souza, N. F. (2020). '(Non)dance and (non)math people': Challenging binary disciplinary identities in education. *Research in Dance Education*, *0*(0), 1-19. https://doi.org/10.1080/14647893.2020.1853692

Lenzi, S., & Ciuccarelli, P. (2020). Intentionality and design in the data sonification of social issues. *Big Data & Society, 7*, 205395172094460. https://doi.org/10.1177/2053951720944603

Leonard, A. E., Bannister, N. A., & D'Souza, N. F. (2020). '(Non)dance and (non)math people': Challenging binary disciplinary identities in education. *Research in Dance Education*, 1–19. https://doi.org/10.1080/14647893.2020.1853692

Lowenfeld, V. (1957). *Creative and mental growth* (3rd ed.). Macmillan.

Malinverni, L., & Pares, N. (2014). Learning of abstract concepts through full-body interaction: A Systematic review. *Educational Technology & Society, 17*(4), 100-116.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung-Byers, A. (2011). *Big data: The Next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Martin, J. (1975). *Introduction to the dance* (New edition of 1939 edition). Dance Horizons Republication.

Moran, G. S., & Alon, U. (2011). Playback theatre and recovery in mental health: Preliminary evidence. *The Arts in Psychotherapy, 38*(5), 318-324. https://doi.org/10.1016/j.aip.2011.09.002

Mumford, M. (2008). *Bertolt Brecht*. Routledge. https://doi.org/10.4324/9780203000991

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. The National Academies Press. https://doi.org/10.17226/11463

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Oliver, J. C., & McNeil, T. (2021). Undergraduate data science degrees emphasize computer science and statistics but fall short in ethics training and domain-specific context. *PeerJ Computer Science*, *7*, e441. https://doi.org/10.7717/peerj-cs.441

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Prado, J. C., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2), 123-134. https://doi.org/10.1515/libri-2013-0010

Papert, S. (1994). *The Children's machine* (Reprint edition). Basic Books.

PARCOS. (2020). *Using Arts-based methods in science communication. LUT University*. https://hackmd.io/@art-based-methods-guidebook/HJMVIhHFL/%2F2UkwoMKeSyugSm59sF2Iew

Perovich, L. J., Cai, P., Guo, A., Zimmerman, K., Paseman, K., Espinoza Silva, D., & Brody, J. (2020). Data clothing and Bigbarchart: Designing physical data reports on indoor pollutants for individuals and communities. *IEEE Computer Graphics and Applications, 41*(1), 87-98. https://doi.org/10.1109/MCG.2020.3025322

Raffaghelli, J. E. (2020). Is Data literacy a catalyst of social justice? A Response from nine data literacy initiatives in higher education. *Education Sciences, 10,* 1-20. https://doi.org/10.3390/educsci10090233

Reynolds, D., & Reason, M. (2012). *Kinesthetic empathy in creative and cultural practices*. Intellect Books.

Ryan, L., Silver, D., Laramee, R. S., & Ebert, D. (2019). Teaching data visualization as a skill. *IEEE Computer Graphics and Applications, 39*(2), 95-103. https://doi.org/10.1109/MCG.2018.2889526

Salas, R., Steele, K., Lin, A., Loe, C., Gauna, L., & Jafar-Nejad, P. (2013). Playback Theatre as a tool to enhance communication in medical education. *Medical Education Online, 18*, 22622. https://doi.org/10.3402/meo.v18i0.22622

Sander, I. (2020). What is critical big data literacy and how can it be implemented? *Internet Policy Review*, *9*(2), 1-22.

Scheaffer, R. L., & Jacobbe, T. (2014). Statistics education in the K-12 Schools of the United States: A Brief history. *Journal of Statistics Education*, *22*(2), 8. https://doi.org/10.1080/10691898.2014.11889705

Solomon, F., Vogelstein, L., Brady, C., Steinberg, R., Thomas, C., Champion, D., Lindberg, L., Enyedy, N., DesPortes, K., Payne, W., Bergner, Y., Taylor, E., & Shapiro, R. B. (2021). *Embodying STEM: Learning at the intersection of Dance and STEM*. In de Vries, E., Hod, Y., & Ahn, J. (Eds.), *Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021* (pp. 819-826). International Society of the Learning Sciences.

Sommer, S. R., & Polman, J. L. (2017). Embodied activities as entry points for science data literacy. In *CSCL 2017 Proceedings* (pp. 849-850). International Society of the Learning Sciences, Inc.

Styan, J. L. (1981). *Modern drama in theory and practice: Volume 3, Expressionism and epic theatre*. Cambridge University Press.

Thorp, J. (2015). *A Sort of joy*. Memo (Random). https://medium.com/memo-random/a-sort-of-joy-1d9d5ff02ac9

Timmermann, C., & Havemann, L. (2020). Critical literacies for a datafied society: Academic development and curriculum design in higher education. *Research in Learning Technology, 28*. https://doi.org/10.25304/rlt.v28.2468

Turkle, S., & Papert, S. (2015). Epistemological pluralism: Styles and voices within the computer culture. *Signs: Journal of Women in Culture and Society*. https://doi.org/10.1086/494648

Tygel, A., & Kirsch, R. (2015). Contributions of Paulo Freire for a critical data literacy. *Proceedings of Web Science 2015 Workshop on Data Literacy* (pp. 318-34). http://www.dataliteracy.eita.org.br/wp-content/uploads/2015/02/Contributions-of-Paulo-Freire-for-a-critical-data-literacy.pdf

UNICEF Rwanda, & National Institute of Statistics Rwanda. (2016). *Teachers' discussion guide for reading data with children*. UNICEF. https://www.unicef.org/rwanda/reports/teachers-discussion-guide-reading-data-children

Williams, S. (2020). *Data action: Using data for public good*. The MIT Press.

Willett, W., Jansen, Y., & Dragicevic, P. (2017). Embedded data representations. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 461-470. https://doi.org/10.1109/TVCG.2016.2598608

Yap, A. C. (2016). *How to teach kids empathy through dance*. The Atlantic. https://www.theatlantic.com/education/archive/2016/01/learning-empathy-through-dance/426498/

Zeidler, D. L., Sadler, T. D., Simmons, M. L., & Howes, E. V. (2005). Beyond STS: A Research-based framework for socioscientific issues education. *Science Education, 89*(3), 357-377. https://doi.org/10.1002/sce

Shapiro, B. R., Meng, A., Rothschild, A., Gilliam, S., Garrett, C., DiSalvo, C., & DiSalvo, B. (2022). “Bettering Data”: The Role of Everyday Language and Visualization in Critical Novice Data Work. *Educational Technology & Society*, *25* (4), 109-125.

“Bettering Data”: The Role of Everyday Language and Visualization in Critical Novice Data Work

Ben Rydal Shapiro[1*], Amanda Meng[2], Annabel Rothschild[2], Sierra Gilliam[1], Cicely Garrett[2], Carl DiSalvo[2] and Betsy DiSalvo[2]

[1]Georgia State University, USA // [2]Georgia Institute of Technology, USA // ben@benrydal.com // am377@gatech.edu // arothschild@gatech.edu // sgilliam7@student.gsu.edu // cicelycmgarrett@gmail.com // cdisalvo@gatech.edu // bdisalvo@cc.gatech.edu

[*]Corresponding author

ABSTRACT: Informed by critical data literacy efforts to promote social justice, this paper uses qualitative methods and data collected during two years of workplace ethnography to characterize the notion of critical novice data work. Specifically, we analyze everyday language used by novice data workers at DataWorks, an organization that trains and employs historically excluded populations to work with community data sets. We also characterize challenges faced by these workers in both cleaning and being critical of data during a project focused on police-community relations. Finally, we highlight novel approaches to visualizing data the workers developed during this project, derived from data cleaning and everyday experience. Findings and discussion highlight the generative power of everyday language and visualization for critical novice data work, as well as challenges and opportunities to foster critical data literacy with novice data workers in the workplace.

Keywords: Data science education, Critical data literacy, Social justice, Data visualization, Workplace ethnography

1. Introduction

> Data collection? … that’s the first time I’ve heard that term – Shannon
>
> I realized that, hey one thing about data, you have to learn it, you have to get in with it, you have to love it for what it is, and you have to be able to love it to organize it. And I was like oh this is too much! – Nia
>
> It’s a trauma there too that keeps me like, “ok keep it easy.” It’s just a cautionary thing that I constantly think about…like do not insult the man in blue…don’t give him a reason to think that we are against the man in blue…like we want to collaborate safely [with this data] and um, not overstep any type of boundaries and also protect them as well. – Leo

These reflections come from young adults at the beginning of their careers in data work, whom we characterize in this paper as “novice data workers.” These people are beginning to learn — or aspire to learn — how to perform data work, but do not have formal training or education in computing, visualization, or statistics. The above quotes represent how these workers characterized early-stage data entry and cleaning work while employed at DataWorks, a workplace training program and data services provider that employs historically excluded populations to work with community data sets. The language they use contrasts with that of professionally trained data scientists, demonstrating how these novice data workers can hold different views of academic terms such as “data collection” and characterize challenges of critical data work differently. Each novice worker also carries personal histories and even trauma, which impacts how they work with data – particularly, in this case, during a project focused on critically engaging with arrest data to support police-community relations.

This paper is grounded in research that identifies everyday language and other sense-making practices as intellectual resources for learning and teaching (Warren et al., 2001; Nasir et al., 2006), and is aligned with critical data literacy efforts that promote justice-centered approaches in computer and data science education (see Benjamin, 2019; Brown, 2021; D’Ignazio & Klein, 2020; Matuk et al., 2020; Pargman et al., 2021; Vakil, 2018; Wilkerson & Polman, 2020; Yadav et al., 2022). We build on these efforts using qualitative methods and data collected during two years of ethnographic fieldwork at DataWorks to characterize the notion of critical novice data work.

We begin by describing our motivation and DataWorks. Then, we review relevant research concerning critical data literacy, situative studies of data science in the workplace, and alternative approaches to visualization as used in the learning sciences. Subsequently, we describe the empirical data and methods that inform our subsequent two-part analysis. First, we present key findings from a thematic analysis of interviews with the first cohort of five novice data workers employed at a DataWorks program we launched at a large public technology university located in the US South. This cohort included three women and two men — all Black and/or African American people aged between 19–26 years old who possessed little previous experience working with data (e.g., typically only a high school diploma or GED) — named Shannon, Nia, Jalen, Leo, and Terrance (pseudonyms). Our thematic analysis identifies their simultaneous unfamiliarity with data science terminology at the start of their employment and ability to characterize and humanize such terminology using everyday language, including gesture.

Second, we use a case study approach to analyze a project in which three of these novice data workers critically engaged with arrest data from their local police department to support police-community relations. We focus on the challenges they faced when cleaning data and engaging critically with data. Likewise, we share the workers' novel use of participatory visualization with our research team to contextualize and humanize arrest data. Notably, the issue of police-community relations is a particularly salient social justice issue for Black Americans in the United States. Police violence remains a leading cause of death for Black men (Edwards et al., 2019); Black youth are stopped, questioned, and physically detained at rates far higher than white youth (Crutchfield et al., 2012; Epp et al., 2014); and tensions between police and the community following recent killings of Black men and women including Michael Brown (age 18), Breonna Taylor (age 26), Tamir Rice (age 12), George Floyd (age 46), Atatiana Jefferson (age 28), and Daunte Wright (age 20) have led to national protests about the racialized nature of police violence. Thus, this case study contributes to studies that have examined personal experiences and narratives of trauma from police violence in Black communities (see Smith Lee & Robinson, 2019).

Findings and discussion synthesize our two-part analysis to highlight the value of everyday language and visualization for critical novice data work, as well as challenges and opportunities to foster critical data literacy with novice data workers in the workplace. We conclude by outlining future research directions that expand on the notion of critical novice data work and address inherent limitations to this qualitative study.

2. Motivation and background

2.1. Broadening and diversifying participation in computing

For decades, United States researchers and practitioners have attempted to broaden participation in computing – particularly for Black Americans – by enlarging the pipeline of potential computer science students. Despite these efforts having been in place since the early 2000s, there has been little measurable impact on the percentage of Black Americans majoring in computing fields. Over the past two decades, for example, only 3.6% of undergraduate degrees in computing have been awarded to Black Americans (Computing Research Association, 2019).

More recently, researchers and practitioners are moving beyond quantitative inclusion of underrepresented groups. They now advocate for more expansive, participatory, and democratic approaches that foreground social justice to foster participation in computing and data science. Some demonstrate how computing education can apply critical traditions in education to interrogate the sociopolitical context of computing education (Vakil, 2018). Others critically explore the role of Black youth in computing and how they have experienced power imbalances (Rankin & Henderson, 2021). Still others use an abolitionist framework to explore Black empowered futures in computing (Benjamin, 2019).

These efforts are beginning to inform a paradigm shift in computer and data science education that motivates our work. This shift encourages researchers to explore race and identity in computer science settings; structure learning environments to address the power imbalance that permeates the tech sector; and conduct critical research on the ethical and fair use of data across school and workplace settings (see Crooks & Currie, 2021; DesPortes et al., 2022; Gray et al., 2022; Jayathirtha et al., 2020, Santo et al., 2019; Silvis et al., 2022; Yadav et al., 2022).

2.2. What is DataWorks?

DataWorks is a new model for providing data services to companies and non-profit organizations in ways that aim to foster diverse approaches to data science, support equitable labor practices, and develop just forms of engagement between universities and communities. DataWorks employees are from historically excluded communities and learn entry-level data science skills such as cleaning, formatting, and labeling datasets (e.g., for use in machine learning algorithms) using real datasets submitted by local companies and non-profit organizations. This training aims to empower workers and support pathways to long-term, full-time employment in other kinds of organizations. Through DataWorks, we hope to outline one kind of model for members of historically excluded communities to learn entry-level data science skills. We also introduce DataWorks to provide an example of a sustained research context and community of practice that develops new knowledge and approaches to teaching data science that acknowledge and respect diverse subjectivities.

3. Literature review

3.1. Critical data literacy

Critical data literacy is vital to broadening and diversifying participation in computing, particularly as we face realities of how technologies perpetuate racism and systemic oppression. Scholars of critical data literacy identify technical and social components of data literacy that are complementary and indivisible (Tygel & Kirsch, 2016). For example, the technical skills required to analyze and work with datasets should be paired with efforts to understand how that data is embedded in specific local contexts, including efforts to work with communities and people who live in those contexts from which the data is about. Additionally, Wolff et al. (2016) posit we are more likely to gain competencies that allow us to learn from and solve problems with data when we participate in a full inquiry process. These scholars echo the work of Bhargava et al. (2015), who assert that data literacy must (1) foster adaptive capacities and resilience instead of teaching platforms and technical languages, and (2) empower people in meaningful and effective ways. Bhargava et al. (2015) also identify a critical challenge to data literacy concerning the need to better understand the importance of context.

To apply and expand on these theoretical foundations of data literacy, we highlight a context in which critique and critical inquiry are dissuaded: the workplace. We also address the challenge of understanding context through analysis of everyday language used by novice data workers and a case study of a project to clean and visualize arrest data, demonstrating how critical inquiry and data work interrelate. Our understanding of the critical data literacy process during this project— whereby data literacy and critical consciousness circulate in a positive feedback loop — comes from Paulo Freire's theory of reflection and praxis in learning in *Pedagogy of the Oppressed* (Freire, 1968). Freire defines how critical consciousness is achieved through reflection in dialogue and practice that involves learning to perceive contradictions. Using the Freirean process of critical consciousness building, we investigate the context of novice data workers as they clean and visualize local arrest data.

3.2. Situative studies of data science in the workplace

In fields including statistics education and computer-supported cooperative work, ethnographic methods have long been employed to study work practices and reveal invisible labor (see Bakker et al., 2006; Noss et al., 2000; Star & Strauss, 1999; Suchman, 1995). Ethnography collects the formal knowledge people use in their work. That is, as a methodology, it examines the people who do the work, not just tools or technical systems for work.

Informed by this history, researchers are beginning to conduct situative studies of data science in the workplace. For example, Muller et al. (2019) use a grounded theory analysis of interviews with professional data scientists to characterize five ways humans influence data work practices and the outputs of data science systems: Through the discovery (e.g., finding a public dataset), capture (e.g., making selections or substitutions), curation (e.g., cleaning or converting), design (e.g., imputing missing or validating data), and creation (e.g., simulating) of data. Passi and Jackson (2018) use ethnographic fieldwork similarly, to reveal how trust is negotiated during the everyday work of corporate data science teams and how this negotiation shapes the development of data science systems. Furthermore, scholars of human-centered data science in the workplace are beginning to ask questions such as: "How might we mitigate the (environmental, social) harms imposed upon workers involved in industrial technology production? How might we design for "good" jobs? How might we as researchers inform policy initiatives that directly influence the conditions of digital labor?" (Fox et al., 2020; also see Aragon et al., 2022).

These situative accounts of data science continue to shape data science education and practice. Yet, they focus almost exclusively on the perspectives of expert data scientists or corporate data science teams (see Rothschild et al., 2022 for an exception). For example, they rarely have considered the experiences of Black Americans in the workplace. In response, we center the perspectives of novice data workers from historically excluded communities in data science. Their perspectives expand on and challenge prevailing accounts and assumptions about data and data work. Further, we detail the types of challenges that can arise for people and organizations through efforts to foster critical data work in the workplace.

3.3. Alternative approaches to data visualization and the learning sciences

Our project case study also draws from a growing body of research in the learning sciences expanding on what we see as alternative approaches to data visualization. While information visualization has prioritized the exploration, analysis, and presentation of raw data (see Card et al., 1999), alternative approaches to visualization foreground alternative ways of knowing, non-binary approaches to data, and the contextual or human dimensions of data (see D'Ignazio & Klein, 2016; Dörk et al., 2013; Lupi, 2017).

The project described in our case study sought to empower workers by using visualization in ways informed by Matuk et al. (2022) who demonstrate how artistic practices can foster an accessible and personally relevant approach to critical data literacy. Specifically, their research draws from Data Humanism (Lupi, 2017) and shows how data-driven art provides ways to contextualize ideas about data science in the real world and pursue personal interests. In our work we provided opportunities for workers to integrate art and more traditional forms of data visualization to embrace ideas such as subjectivity and serendipity over objectivity and prediction and to encourage more personalized visualization designs and grammars. Similarly, our understanding of data cleaning and visualization as both technical and social processes is informed by Kahn (2020), who highlights social and familial dimensions of technical practices such as data wrangling. Notably, her work leverages new digital mapping and dynamic geovisualization tools that allow youth and families to link personal reflections about their own data with broader societal issues. These broad issues are represented in aggregate data through interactive, digital maps that represent personal, family migration stories or family geobiographies (also see Taylor & Hall, 2013; Shapiro et al., 2020). Finally, we draw from learning scientists demonstrating the need for data visualization pedagogies to better support teachers and students to negotiate racialized contexts of data as they emerge during discussion (Philip et al., 2016). These ideas begin to outline aspects of a critical visualization pedagogy that require an equal commitment to develop racial literacy through environments that interrogate processes of race, racism, and racialization.

We draw from and contribute to this body of work by exploring everyday perspectives novice data workers use to understand data visualizations and illustrating new roles visualization can play to support critical novice data work.

4. Methodology

We answer the following research questions:

- What sense-making resources (e.g., ways of using language, experience, social practices) do novice data workers bring to data work?
- What challenges do novice data workers face as they begin to critically engage with data for the first time?
- How does the process of cleaning data support novice data workers' critical questioning and visualization of data?

Our analysis is broken into two parts. First, we share a thematic analysis of semi-structured interviews with Shannon, Nia, Jalen, Leo, and Terrance as they began their employment at DataWorks. These in-person interviews lasted 1–2 hours and elicited workers' perceptions about data and data work through questions including: What does the term data mean to you? Where or do you encounter data in your daily life? Can you describe any representations of data you encounter in your daily life? We analyzed their responses using thematic analysis in the grounded theory tradition (Charmaz, 2006; Glaser & Strauss, 1967). All interviews were video recorded, and transcripts were produced by the research team. Afterwards, we conducted an open coding process, identifying themes relevant to how workers perceived data and data work. We then iteratively discussed and refined these themes over five months to produce the key themes presented in this paper.

Second, we use a case study approach (Yin, 2009) to analyze audio and video recordings, representations, and reflections collected during a real world, 10-week long project completed by three of these workers at DataWorks approximately one year into their employment. Our motivation for this project was multifaceted.

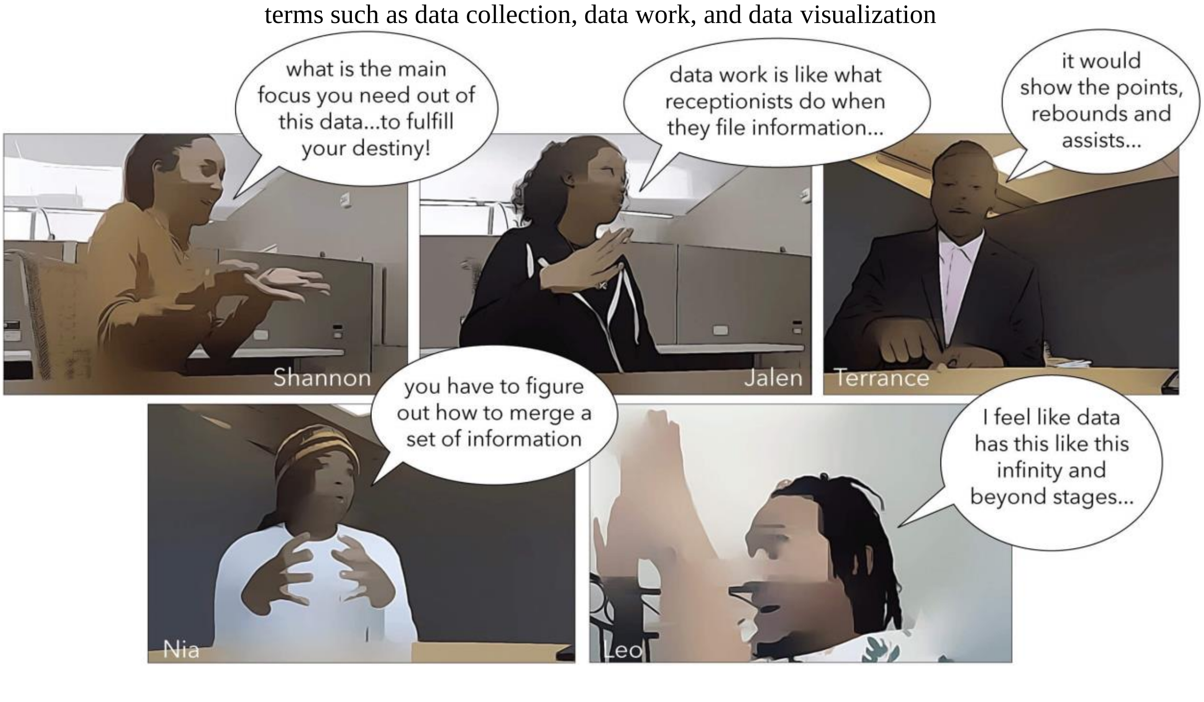
First, this project was the first time that workers developed their own data cleaning plan, as opposed to entering data, allowing us to see how novices approached such work for the first time. Second, most projects at DataWorks required highly specific deliverables for clients and provided little time for critical or creative dialogue and reflection. DataWork "clients" were members of a local criminology department who provided datasets and orienting tasks for the project, which empowered us to critically engage with and communicate arrest data in ways that foregrounded workers' perspectives on police-community relations. Finally, as we opened this paper, the issue of police-community relations is a particularly salient social justice issue for Black Americans.

Before continuing our analysis, we acknowledge the positionality of our research team, all of whom are from the United States. The lead author, who interacted most with workers in the workshop, is a white male faculty member whose research draws from the learning sciences and critical visualization. The second author is a white female faculty member whose research focuses on data activism. Other authors include a Black female who is a leading racial equity consultant in the city where DataWorks is located, and two white faculty members, one male and one female whose research spans participatory design and learning. Additional team members include a white Ph.D. student and a Black Ph.D. student who have educational backgrounds in computer science. Team members have a long history of working with non-profit and civic organizations in the region and have consistently participated in outreach and community-based research with the goal of eliminating structural forms of oppression and to distribute the resources and opportunities we have access to for the benefit of minoritized communities. However, we acknowledge that we are not positioned to contribute to scholarship on the Black American experience. Rather, we see the context of DataWorks as an opportunity to better understand the development of a democratic data workplace that is inclusive of individuals from excluded groups and to better understand how and why our concepts of data need to be expanded to account for alternative perspectives in computer and data science. In this way, the experiences reported in this paper with lower income and Black novice data workers provide insights into our income privilege and whiteness and how that impacts data work.

5. Thematic analysis of intake interviews

Figure 1 shows excerpts from semi-structured interviews with each worker as they began their employment at DataWorks. In the following, we unpack these excerpts and share three primary themes that emerged from our analysis of these interviews.

Figure 1. Excerpts from interviews with each novice data worker at the start of their employment describing terms such as data collection, data work, and data visualization



5.1. Formal data science terminology is unfamiliar

Professional data scientists, researchers, and educators frequently assume that terms such as data, data collection, and data analysis are familiar to a general audience and students. An initial finding from our analysis of interviews with each novice data worker suggests this is not always the case, even for those who are aspiring to work with data in their professional careers. Two workers indicated the terms data and data collection were entirely new to them. For example, when asked about the term data collection, Shannon responded, *"Data collection? … that's the first time I've heard that term."* While each of the other three workers shared that they had heard these terms before, they also indicated they were unsure of their meaning. For instance, Terrance defined data as *"The everyday use that people use in the world to like know what's this and like what would be this and how would we be able to use it to make the world a better place."* Further, all five workers we interviewed indicated that terms such as data analysis and data exploration were completely new to them.

5.2. Everyday language to humanize and contextualize data science terminology and practices

While these terms may have been new to workers at first, with some prompting by the interviewer to explain these terms in relation to their own lives, each novice data worker was able to characterize them using everyday language. Their everyday language highlighted what we interpreted as human and contextual dimensions.

For example, when asked to define the term data collection, Terrance shared *"collecting raw data and using it."* Terrance was then prompted to describe the term in relation to basketball, a sport he strongly identified with, and expanded his definition through an example saying, *"First, I would show, like, it would be a name and it would show the points, rebounds and assists and, like, every point that he makes you write down, was it a two pointer or was it a three pointer or if it was a rebound, like, how many rebounds did he grab. You write, tally it down, and it would be a total. You sum that up, and you put that down on the sheet."* While making this statement, Terrance gestured as if he was constructing a data table with each row characterizing a basketball player and columns indicating rebounds, points, and assists. The top right image in Figure 1 shows Terrance using a gesture to construct one row of this table.

Similarly, in characterizing the term data work Shannon shared, *"I guess what it means is just like when you're working with that data it's like you are trying to figure out the overall like what can you do to better that data I guess um I'm not cuz I don't really work with data so once I do I feel like the overall when I do work with it it's just to better that data."* Subsequently, she posed different questions one might ask about data or data collection including *"What is the best way to collect data in the best way?"* These questions in turn led her to describe concepts such as the notion of *"fulfilling destiny."* Fulfilling destiny characterizes the primary goal of data work in Shannon's view: to make data useful to other people and organizations — that is, to fulfill the data's destiny. The top left image of Figure 1 shows the moment Shannon used the term fulfilling destiny. It shows her gesturing as if she is handing a data set that she has made useful to another human being or organization, to describe fulfilling their destiny. Informed by Kahn (2020) as well as Van Wart et al. (2020), we suggest Shannon uses everyday language and gesture to characterize aspects of data wrangling as an inherently social and humanistic practice in ways that expand dominant conceptions of data wrangling (also see Rubel et al., 2017).

While Leo also at first struggled to talk about the term data, he eventually defined data as *"Yea, like music gives a lot of detail about you know where you are in your life. I feel like, you know, 24-25 [year-olds] most likely are going to listen to something a little bit more inappropriate…but then again you have to think about what year they were born in…and to me, that's, like, data…to me, you gotta put the year they were born in to see what genre of music was out there, uh, collecting it from people you know and see what it is."* Leo, who sang in his school's chorus for many years, uses everyday language and music to characterize how age and genre of music must be understood relationally to figure out "where you are in your life." We suggest his descriptions resemble the practice of data contextualization, or the notion that data must always be understood in context.

5.3. Using gesture to communicate data science terminology and practices

Gesture is a fundamental resource in human communication and has been studied in many disciplines including mathematics and science education (Alibali & Nathan, 2012; McNeil, 1992). Yet, computer and data science education research rarely studies gesture. Extant research primarily focuses on how teachers use gestures to communicate computer science concepts to students in classrooms (Solomon, 2021) rather than how gesture is culturally organized and improvised in situated activity (Rogoff, 2003; also see Davis et al., 2020).

Gesture was used by all data workers throughout their interviews to communicate data science terminology and practices in ways they could not do through talk alone. The previous analysis of Figure 1 highlighted Shannon's use of gesture to characterize the concept of fulfilling destiny (top left) as well as Terrance's use of gesture to construct a representation of one row of a data table (top right). Figure 1 also shows a moment when Jalen used a gesture to help her make an analogy between a receptionist who organizes files to a data worker who organizes data. Jalen's hand represents a file that she is slowly placing into a file cabinet as a receptionist would. Likewise, the figure shows Nia using a gesture to describe data analysis (bottom left). During the pictured moment, each of her hands represents a different set of information, and she brings her hands together to represent the merging of information sets necessitated by data analysis. Finally, the figure shows how Leo (bottom right) is using a gesture while he makes the following statement: *"So, I feel like data has like this infinity and beyond stages where you can do thousands of steps to complete whatever the data project that you are doing."* We suggest he uses gesture to characterize time and particularly, the significant amount of time he feels it can take to complete data-oriented projects.

These excerpts show the varying ways workers used gesture to amplify the concept of fulfilling destiny; make an analogy between a data worker and a receptionist; construct a representation of a data table; describe technical aspects of merging data sets; and characterize time through the notion of infinity and beyond.

In summary, we identified three themes that emerged from our analysis of interviews with workers as they began their employment: The unfamiliarity of data science terminology, use of everyday language to contextualize data science practices, and use of gesture to communicate data science terminology and practices. Our following case study analysis presents a different kind of analysis that expands on some of these themes and particularly, the role of everyday language in critical novice data work.

6. Case study analysis

In the second part of our analysis, we describe Nia, Leo, and Jalen's experiences during a 10-week project to explore using data to support police-community relations. The project came to DataWorks from academic colleagues who work in a criminology department. In considering the project, our motivation to involve these data workers was three-fold: Empower workers to critically discuss policing through data collected in their local communities while simultaneously challenging and expanding our own worldviews and value systems about policing; engage workers in the process of creating their own data cleaning plan; and create data visualizations informed by workers interactions with police and oppression experienced by the communities they lived.

We first describe an initial workshop that built on our existing relationships with these data workers and aimed to empower workers to decide if they wanted to take on the project. Subsequently, we analyze their progression over six weeks to develop and carry out a data cleaning plan for an arrest dataset. Finally, we illustrate how workers visualized and critically reinterpreted this data in collaboration with our research team.

6.1. Introductory workshop

The workshop opened with an hour-long conversation about workers' alignments to arrest data and policing. Workers indicated they had never seen arrest data before and were deeply interested in this project because of personal experiences with policing. Leo and Nia willingly shared they had multiple encounters with the police and had been arrested multiple times. One described the dehumanizing experience of being *"locked up for 33 days"* and unable *"to shower for 5 days."* They also shared positive relationships they had with police; one worker described enjoying spending time with an officer in their neighborhood regularly. Among all the workers, police had a significant presence in their lives and communities.

After the discussion covering workers' personal, racialized, and political experiences with policing, we visually explored a sample tabular arrest dataset from the client that encoded arrests, along with officer activity from a local police department from 2014–2019. Arrests in the dataset were encoded with spatial features such as address, longitude, and latitude; temporal features such as the date and time; classifications based on gender, race, and age; and standardized measures for reporting crime used by police departments through systems such as the Unified Crime Reporting (UCR) Program including police beat, officer shift, and arrest type (e.g., homicide, burglary, theft).

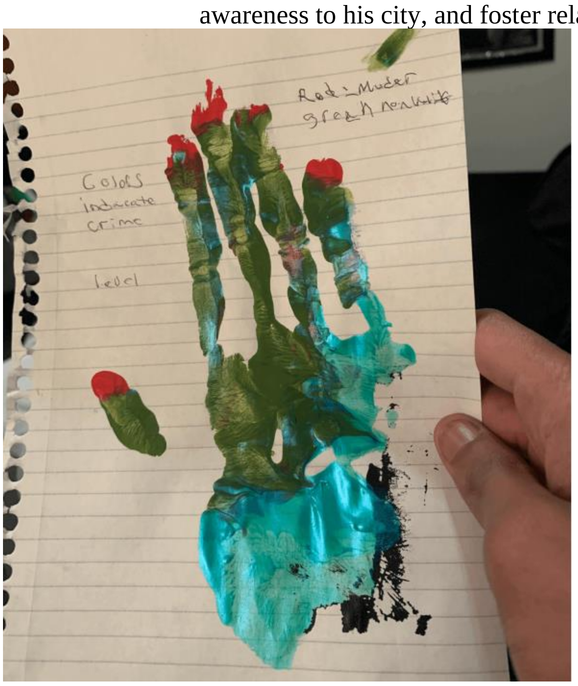
We then employed visualization tools to explore this data. While workers found such tools "cool," they struggled to see their value and felt viewing the data as a table was more meaningful because it was a familiar representation and yielded more details of the data. Later in the analysis we return to their evolving perceptions about visualization.

The table view allowed each worker to recognize the enormous racial disparity in reported arrests as well as specific places encoded in the dataset. As Leo described, *"We know all… I know every place on this right here."* The first hundred rows of the dataset predominantly displayed arrests of people encoded as Black from neighborhoods in which the workers lived and worked. Their lived experiences with police also led them to ask and answer questions about what constituted an arrest, what information went into a police report, and how other people and officers interpreted this dataset. Likewise, they were able to see how data become encoded from their lived experiences in ways that were new and previously invisible to them. The workers also suggested their own names for column headers such as UCR_Literal (a standard way to characterize arrest type), which signaled the type of arrest made by an officer. In this case they suggested *"gotcha"* for the header name.

As we collectively viewed this dataset, the research team introduced and discussed the notion of unreported arrests, inherent bias in arrest data collected by police departments; how the use of predictive policing systems based on such data reinforced the marginalization of communities where workers lived; and other critical questions about arrest data. These topics were quite new to workers and became areas we would revisit throughout the project.

At the workshop's conclusion, workers emphasized they wanted to take on this project. To further empower them to shape the project in ways that were uniquely encouraged by the client, we asked them to reflect on the following prompt: What ways would you like to use these data with your community, family, or others as part of this project? Figure 2 illustrates part of Leo's response that became central to later stages of the project. Leo used paint to create a handprint of his hand where color encodes different arrest types and the amount of space represented by each color on the handprint corresponds to Leo's rough interpretation of the number of arrests for different arrest types. For example, red encodes homicide and covers only the fingertips of each finger indicating fewer homicides that occur in comparison to other arrest types in the dataset. In the figure text, Leo explains his view of the handprint as a method to humanize arrest data and speak both to police and the community.

Figure 2. Leo's initial sketch and written reflection excerpt of a handprint to represent arrest data, bring awareness to his city, and foster relations between police and the community



"I feel this will bring the most awareness to our city. I chose these because of the reactions I received from family friends and a few strangers yesterday from examples. Far as being able to connect too on a human to human Level when looked at and understood. The handprint Is Definitely Something That shifts a person perception and allows them to bond with our crime data."

6.2. Data cleaning

Subsequently, workers developed and carried out their own plan to clean a publicly available 2020 arrest dataset from a local police department. We selected this dataset in collaboration with our client because it was formatted using two different crime reporting systems, coinciding with the start of a national effort to update ways arrest data were encoded. The data merging project opened space for workers to engage in common and important cleaning tasks including merging two differently formatted data tables, standardizing column formats (e.g., standardizing date stamps), and addressing missing data and encoding errors.

For each worker, developing an initial set of steps to begin cleaning this dataset was an imposing new task, one they viewed as extremely challenging and highly creative. For example, Nia and Leo shared the following exchange after attempting to develop an initial set of steps:

Nia: I'm going to be honest I have no structure with this data…I don't know what I'm looking for, I need some instructions…I feel like I'm looking for it myself and I don't know…I'm so used to us running the play

Leo: If I'm a doctor but you just gave me a quick little prep sheet and I'm running in here and then you got his whole body showing up and you just showing me his foot and you tell me to untie something...I'm not going to know the full extent of what is going on up all the way up under

Nia: Yeah there's no hand walking…if I don't get this step right…he's dead! Laughing…that's how I felt with this…If I don't get this one thing right it's not machine readable

Leo: Yeah but you see we have the freedom we got the freedom that's what they telling us…that's what we got to keep in mind we have the freedom to go about it like we're free right now that's why it's hard

As Nia describes, in prior projects the workers focused on "running the play" by following clear, predefined instructions about how to input and format data. Leo extends Nia's statement, analogizing the challenges of developing initial cleaning steps to a doctor being able to see the entirety of a patient's needs, but only having information about that patient's foot. As Leo summarizes, having the "freedom" to develop their own set of cleaning steps was new and challenging.

Figure 3. Diagrams and README file excerpts created by workers during data cleaning

A

Look through Data and list possible questions for Client, meet with Client to Identify those issues.
Possibly look at old data sets on the website.
List out everything we notice in the dataset from previous projects to help gain a better understanding.
Figuring out police jargon and codes.
Compare other police datasets from different states, and cities.
Meet with Client to ask how they want the Data and to clarify issues that we see in the data.
Define Headers and make them more understandable using the Headers from the old data set.
Look through Column Headers from New and Old Data sets to see if they match.
Combine Data Sets Old and New Adding in New Columns if Necessary.
- We can figure out how many rows are within the dataset.
- Look through and discuss why the dataset is messy or why it needs to be cleaned.

Remove Columns if necessary.
Custom Sort each column and identify other issues in the Data
Figure out which header we will filter to start the cleaning process.
- We could use date, location, npu, and charges to start off the cleaning process.
- Make sure everything looks the same such font, format.
- Use Google Maps to find the exact location using longitude and latitude coordinates.
- Ask how each column helps the next. (Npu, Neighborhoods)

Make those changes and corrections in the Data Set.
Present our final spreadsheet, gain feedback from client, and make last corrections before sending it off.

B

PREP:
1. Look through Data and list possible questions for Client (e.g., police jargon and codes)
2. Look at one really good example of a cleaned crime dataset
3. List out everything we notice in the dataset from previous projects to help gain a better understanding.

CLEANING STEPS:
1. Combine 2 datasets from 2020 *(January-August, September-December files)*
    1. *Compare columns*
    2. *Decide which columns can and can't be "merged"*
    3. *Check for most critical errors (e.g., N.A., blank column)*
2. Custom Sort each column and identify other issues in the Data (like housing project)
    1. Decide if you can correct any identified issues/errors or note errors for client
3. Document how much missing data or errors are in the combined dataset *(e.g., highlight, make new tab, OR provide exact counts of # of error rows/columns)*
4. Merge 'occur_time' and 'occur_date' columns into single column called "time"
5. Order by new "time" column
6. Final formatting/address missing data
7. Present our final spreadsheet, gain feedback from client prior to visualization/processing

C

Steps	Description[1]	Example(s)
1	1. Combine 2 datasets from 2020 *(January-August, September-December files)* a. Compare columns. b. Decide which columns can and cannot be "merged" c. Check for most critical errors (e.g., N.A., blank column) d. Hid, and changed name to the columns (e.g., Shift to Watch)	Copy January –August and paste it into September – December or vice versa Report # | Report Date | Occur Date | Occur Time | Possible Date Report # | Report Date | Occur Date | Occur Time | Possible Date
2	2. Custom Sort each column and identify other issues in the Data (like housing project) To find these quicker you must Create a new column at the far left hand corner of the screen. Then you to the end of the last row of the data sheet and write in the add column Null value, Then under that Blanks,NA ONCE That is done apply =CountIf(range,critera) in the the Column next to null . Fill in the range and Value inside the Formula. Do The same the For NA as you done null. Then you use the =CountBlank(formula To receive the Blanks in within. Please Note: use correct Excel literature To get correct results from formula. THEN GET YOUR TOTAL By =SUM(aDDINGT HE PASS THREE FORUMLAS. **a.** Decide if you can correct any identified issues/errors or note errors for client.	Null | =COUNTIF(Null | 0 NA | =COUNTIF(NA | 0 BLANKS | =COUNTBLANKS(| 0 TOTAL | =SUM(| 0 Be sure to add correct conditions in formula!
3	3. Document how much missing data or errors are in the combined dataset.	(**e.g.**, highlight, make new tab, OR provide exact counts or # of error rows/columns)
4	4. Merge 'occur time' and 'occur date' columns into single column called "time". • Practice Sheet	e.g., 'Occur time' + 'Occur date' = 'time' =Sum(A2, B2)
5	5. Order by new "time" column and address missing Data. • Look up Missing NPU Data. • Look up missing Longitude and Latitude. • Try to fill in as many NULL values as possible.	Blank | Blank | Blank | Blank N/A | NA | N/A | N/A Sort in numeral order and filter for missing information
6	Present our final spreadsheet, gain feedback from client prior to visualization/processing.	Power point, word doc.or Exel spreadsheet

Figure 3 shows representations workers created at the beginning and end of their cleaning process. Figure 3A is a list created by workers outlining initial steps they created to clean this dataset. On one hand, it shows how workers were able to develop several clear steps important to data cleaning for this project, including recognizing the need to merge columns from different data tables and drawing from past projects to inform tasks for this project. On the other hand, it highlights a key challenge workers faced regarding how to determine what level of detail should be captured in each step. For example, lines such as "We could use date, location, npu [Neighborhood Planning Units], and charges to start off the cleaning process" are abstract. As workers would find, such steps were not detailed enough to characterize specific tasks required for the dataset. Figure 3B illustrates a subsequent list workers created a few days later that refines cleaning steps into "prep work" and "cleaning tasks." This list is more descriptive in its overarching organization and ordering of tasks, highlighting workers' ability to refine their initial steps. For example, they recognized that they needed to check data formats for each column to merge columns from different data tables.

Figure 3C was produced at the end of their cleaning process. It is an excerpt of workers' attempts to translate their steps into a README template file provided by our client. README files are used to describe a dataset's contents, summarize data cleaning tasks, and provide other contextual documentation for future self and outsider comprehension. Such documentation is essential to conducting efficient, useful, and ethical data cleaning work. We introduced workers to practices of creating README files as they began cleaning data and workers iteratively updated a collaborative README file each week as they cleaned data.

The figure highlights an innovation workers developed to support their collaborative development of a README file. Specifically, they used color-coded text to indicate which worker performed certain tasks, to better support their collaborative work. At the same time, the figure helps reveal several challenges the workers faced when translating more informal descriptions of cleaning steps as represented by lists into a more formal README file representation. For example, providing a specific example of how they performed tasks for each step proved extremely challenging. Likewise, while indicating which worker performed each task was important for their own data cleaning work, this information would be less useful to others who would read and possibly use their README file in the future. Moreover, workers focused on carrying out their initial steps in the README file, even if they discovered that adding or removing other steps would enhance the data cleaning process. In other words, changing their initial steps was seen as too challenging. Thus, the same initial six steps from Figure 3B are reflected in Figure 3C.

In summary, this figure illustrates these data workers could develop a data cleaning plan for the first time, but also documents challenges they faced. We discuss later how many of these challenges highlight opportunities to expand the design of traditional data cleaning tools and practices to better support novice data work.

6.3. Participatory visualization

The final stage of the project began with a two-hour workshop where workers imported and explored their cleaned dataset with different exploratory visualization tools including a geovisualization tool we developed in prior work called the Interaction Geography Slicer or IGS (see Shapiro & Pearman, 2017). This workshop also served to assess the quality of their cleaned dataset. During this activity, workers engaged with interactive visualizations in ways quite different to the start of this project when they favored exploring data through tabular representations.

Notably, workers realized and focused on the value of interactive visualization to reveal what they identified as errors in the data and errors in their data cleaning process. For example, workers particularly engaged with geovisualization tools that provided ways to selectively highlight and understand how many geocoded arrest types with null values were clustered around locations such as an airport. Likewise, they were able to identify dates that had been incorrectly merged during their cleaning process. These experiences led Leo to state, "*This is actually a great way of like once you form like a good visualization like inputting your data and working between the spreadsheet and that would be so much easier especially visualizing it and not having to like do extreme amounts of research. …that's pretty cool…that's pretty dope."* In this statement, Leo identifies there can be "good" visualizations better suited for particular datasets, and he highlights the value of going back and forth between cleaning and visualizing data to identify errors in a dataset to better inform a data cleaning process. He contrasts this process with doing "extreme amounts of research" to try to identify errors from a tabular representation as he and the workers had done during their data cleaning process.

Workers also recognized the value of color and interaction techniques such as sorting or linking different views in a visualization to see trends in the quantity and type of arrests across different neighborhoods that were challenging to see in a tabular representation. Many of workers' questions focused on comparing the type and number of arrests across different neighborhoods and neighborhood planning units. Likewise, while doing such comparative analysis, workers repeatedly made statements such as "this makes it real." We interpreted these statements as further evidence that they found visualization more meaningful than before because they gained a sense of ownership of the dataset from developing and carrying out their own data cleaning plan.

To conclude this workshop, workers sketched different visualization ideas on post-it notes as shown in Figure 4A. The color of the post-it indicates the worker who developed the idea. These ideas included traditional bar charts and scatterplots, virtual reality experiences, and representations integrating digital maps and handprints through tools the workers had been exposed to including the IGS.

Figure 4. Initial exploratory visualizations of arrest data created by workers

Figure 4B-D shows more refined visualizations that workers developed on their own time during the week following the workshop. For example, Figure 4C shows a series of drawings by Jalen, reflecting approaches to visualizing data she had learned previously in school. These drawings include bar charts grouped by a unit of time (e.g., a year or month) with color encoding type of arrest to compare how arrests varied over different time periods. These drawings also included the beginning of a histogram (middle drawing), which Jalen described as *"more specific"* than bar charts. Notably, her histogram only begins to sketch out potential ranges to group data. Leo developed quite different visualizations including a unique badging system for police officers, as well as a three-dimensional hand created by Leo that furthered his initial notion of a handprint. These visualizations by Leo became meaningful ways to see and talk about workers' critical perspectives on this data at a subsequent project discussion during which Leo made the following statement to explain his officer badging system.

I've been around police on the wrong end and presenting something that tracks their badges doesn't really sound like…it sounds kinda risky to approach with their own data…You know how we think about police you know don't give them a reason to think that we're trying to get them individually...I thought collectively as a county as a whole when they work together that wouldn't be as irritating and would be more acceptable instead of getting them like physically badge by badge.

Leo's statement reveals two things. First, it highlights a personal approach to developing police officer badges. This approach displays arrest data on a badge at a collective or county level as opposed to an individual or officer level, to protect officers' privacy and foster relations between the police and community. Second, his statement shows how the personal experiences and trauma he experienced through encounters with police shaped his critical perspectives. Put simply, exploratory visualization allowed Leo and other workers to express that they were afraid of criticizing police officers, due to fear of police retaliation but also due to what we interpreted as a genuine value for collective action with and not against police officers.

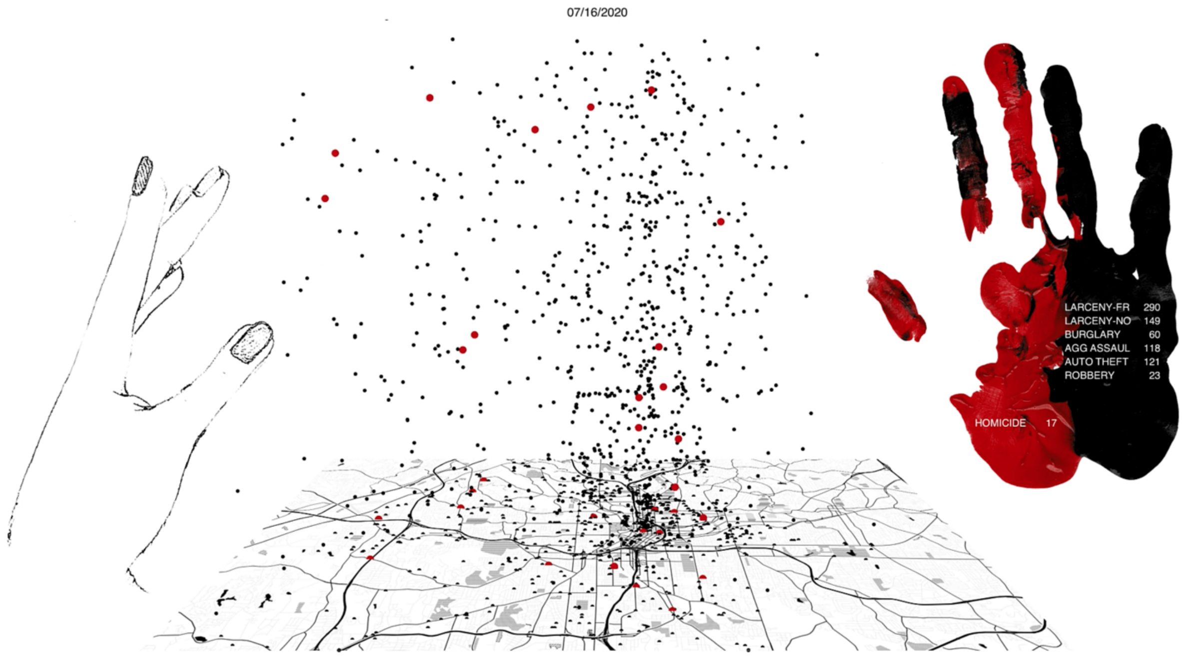
From these many ideas, the workers decided to focus on integrating Leo's notion of what workers began calling *"the crime hand"* with digital maps and arrest data visually animated in the IGS. Jalen described how this approach provided a way to *"humanize"* the data; Nia and Leo described this approach provided ways to communicate to the police and community; and Leo further emphasized this approach reflected ideas such as "a hand holding the city's problems" and experiences from encounters with police such as having their fingerprints taken from them when arrested.

To advance this idea, Jalen and Leo developed many sketches of their own hands, some of which are shown in Figure 5. These sketches also explored coloring schemes workers felt were most meaningful to the communities they lived and different hand positions that could interact with data displayed in the IGS.

In a final workshop together, we integrated these hands with interactive IGS visualizations. This necessitated demonstrating to workers how to use tools such as Adobe Photoshop to cut out their crime hands and a video editor to place them as shape files over interactive IGS visualizations we recorded together. This process generated several decisions and critical conversations.

Figure 5. Crime hand sketches by Jalen and Leo

Figure 6. Screenshot from final video produced by workers and the research team integrating crime hands with IGS visualizations of arrest data (Video available at https://youtu.be/RROp5xg59gw)



For example, workers first created data visualizations with the IGS that used several colors to encode arrests based on neighborhood planning units. This allowed comparisons of the amount and type of arrests made by police across different neighborhoods and how these evolved across 2020. Yet workers felt the use of multiple colors did not convey the dark and more serious tone they wanted to communicate. Thus, they decided to use only two colors in their final visualization: Red to encode homicides and black to encode all other arrest types. Likewise, we had critical discussions about workers' and our own stances regarding arrest data and policing. These discussions revisited previous comments and assumptions shared by workers and informed decisions about data encodings, views, and the goals of a final collaborative visualization. Notably, our research team expanded upon issues such as bias in arrest data collected by police and predictive policing software to share how such work inspired us to pursue a final visualization critical of certain policing tactics and such uses of arrest data. Yet, workers continued to share they did not want to make something that could be interpreted as being critical of police officers. For example, as a potential title for the final visualization workers suggested: *"There is a thin line between crime and mistakes, the choices go hand in hand."* Through such titles workers suggested orientations where they and their communities assumed blame for past encounters with officers.

Throughout this project, but especially during these conversations, the workers and our research team experienced contradictions, tensions, and risk that were deeply meaningful and expanded both our worldviews and understandings of data. For example, members of our research team who had not experienced trauma and fear of an arrest at the hands of police, were able to experience white privilege and how our relationships with the data were distinctly different from the data workers. Likewise, workers began to appreciate bias and power dynamics encoded in arrest data collected by police departments such as the notion of unreported arrests. Along with a desire to protect the privacy of individuals encoded in the data, an awareness of power inspired the workers to create a visualization that presented an aggregate depiction of arrests across their entire city as opposed to focusing on a few neighborhoods where they lived.

Figure 6 is a screenshot from a final video-based visualization that reflects these experiences and negotiations. It aims to encourage more contextual data-driven conversations about arrest data to support police-community relations. This video can be viewed at the following link: https://youtu.be/RROp5xg59gw

7. Discussion

7.1. The generative power of everyday language and visualization for data science education and practice

Data science is an emerging discipline in need of more expansive language and terminology to expound, teach about, and communicate the contextual and human dimensions of data. This paper underscores the generative power of novices' everyday language to contextualize and humanize data science concepts and terminology. For example, terms such as "bettering data" and "fulfilling destiny" coined by workers in this paper meaningfully characterize humanistic dimensions of technical concepts such as data wrangling. Our thematic analysis also illustrates how everyday language encompasses non-verbal dimensions, and that recognizing the multimodal nature of everyday language in data science is important. In particular, gesture provided a way for novices to draw from their own cultural and historical backgrounds in communicating more formal and unfamiliar data science terminology and practices. Likewise, across the project described in this paper workers drew from their personal and cultural experience with policing to describe arrest data and challenges associated with cleaning and being critical with data. We suggest their language has a descriptive power that is missing from current language used in data science. It would be valuable, for example, to develop more varied and diverse concepts to structure and work with data from the perspectives of the places and communities a dataset describes (see Eglash et al., 2021).

This paper also illustrates the value of incorporating Black novice data workers in data visualization. Their decision to include the hand and handprint to communicate and creatively reinterpret data was done to humanize the data. It prompted the viewer to connect, even identify with and embody the data pouring down onto the map in the IGS. Workers' invention and development of the crime hand reflected their connection to crime formed by their direct experience with policing. The crime hand served as a novel and meaningful approach to communicate more than data to diverse audiences. The hand, and the activity of designing the data visualization served as another form of dialogue and communication between the data workers and our research team.

The final visualization presents an artifact that we suggest is valuable to emerging discussions and perspectives in data science education and practice. We find Black feminist theory and particularly, the theories of Lorde (2007) and hooks (1988), useful to understand and interpret this data visualization. Notably, Lorde and hooks make it clear that embodied emotion and intuition must not be ignored to gain knowledge and ultimately achieve liberation. With this theoretical lens we can begin to understand the data visualization as a product of a deep knowing and of an ongoing articulation of arrest data and the experience of being Black in a police state. Data workers had to both understand and articulate the arrest data and the experience of arrests to produce a visualization that embodied both. Likewise, in our analysis we highlighted how some of the workers' decisions in creating this final visualization illustrated forms of critical data literacy. They chose to communicate emotion through more aggregated depictions of data due to potential biases encoded in arrest data and to protect the privacy of individuals represented in the data.

Our interpretation suggests the final visualization contributes to scholarship highlighting how alternative visualization approaches that foreground more personal and emotional aspects of data support critical data literacy efforts and pedagogies (Matuk et al., 2022). However, one could also argue that the final visualization is evidence that the emotional impact of the data on this project overwhelmed workers' and our own interest in analysis, resulting in a visualization process and final visualization that focused too strongly on communicating emotions as opposed to insights from the data. Alternatively, one could also argue that the representational system of the space-time cube on which the IGS is based provides powerful ways to communicate emotion and tell stories with data but backgrounds more traditional forms of exploratory data analysis that are important for novice data workers to engage with as they begin to work with data for the first time.

7.2. Fostering critical data literacy with novice data workers in the workplace

Throughout the project described in this paper we drew on existing research to introduce and scaffold critical perspectives of policing and arrest data. We hoped such an approach would empower novice data workers and challenge our own worldviews and value systems. Reflection and dialogue comprised our primary methods of engaging novice data workers in critical conscious building during group meetings. While forms of oppression and trauma that workers experienced through past encounters with police fostered deep interest in this project, they also caused triggered conflicting feelings. They continually expressed that they did not want to be critical of police for fear of retaliation, even as they held a genuine respect for officers. The participatory visualization stage of the project rendered this tension especially visible. Whereas we supported critique of policing and arrest data, what emerged was data workers highlighting the humans in the data, the concept of intentionality, and an invitation to bond with what the data represented. But their work was not completely absent of critique. Inserting their own hand and handprint into a data visualization critiqued how to consume the data, or at the very least, was an invitation to embody the data. Likewise, their choice to create more aggregated data visualizations of homicide and all other arrest types served to obfuscate and thereby protect the humans the data represented. Such choices also align with contemporary social movements (e.g., Data for Black Lives, Abolish Big Data) that advocate for abolishing "big data" collection of Black bodies. Without training in forms of data protection and data privacy, the workers intuitively used such a method to protect and to reduce harm of visualizing arrest data.

These findings highlight two important realities specific to our workplace case study with novice workers. First, when we conceptualize critical data literacy in the workplace, we must account for how dominant conditions of work can obscure the very histories and perspectives a critical data literacy is meant to foreground (see Johnson et al., 2021). Second, Freire highlights the importance of trusting your ability to reason, and that whoever lacks this trust will fail to initiate the dialogue, reflection, and communication that leads to critical conscious building and liberatory action. Freire would explain that this first step towards self-affirmation or self-efficacy precedes the choice of critique. Those who lack self-affirmation or self-efficacy will choose the security of fear over the risk of freedom. With this theoretical grounding, we recognize that being critical with data is more accessible to data workers with a sense of self-efficacy. However, with this project we challenged both workers' capacity to critique and their capacity to design and execute a data cleaning process, which they struggled with. After the conclusion of the project, however, we did see the workers begin to self-affirm. As Nia explained, *"The crime data project ALONE has definitely pushed us and our mindset in a different space and aspect because this is a project that we had to run, like, completely from the beginning to the end… this was something that actually made us who we are today and communicate our…you know what we do today."* It is not that critiquing data is a privilege, but that critiquing data comes from a place of power and a place of knowing. In this context, self-efficacy to clean and visualize the data was as important for generating critique as the situated knowledge the data workers brought from their lived experience.

8. Conclusion

In this paper we characterized the notion of critical novice data work, highlighting the generative power of everyday language and visualization for that work. We also presented challenges and opportunities to foster critical data literacy with novice data workers in the workplace. We conclude by outlining several research areas to expand on this work and build on inherent limitations of this qualitative study. First, this paper highlights the need to study novice data work in different kinds of workplace settings. Such research offers opportunities to contribute more descriptive language about the human and contextual dimensions of data and data work to inform data science and data science education. Second, our work highlights challenges associated with supporting critical data work in the workplace. Put simply, our work raises questions to orient future critical studies in the workplace such as how do we teach critical approaches to data and data work while also sustaining a business practice? Third, this paper begins to detail the types of challenges novices face when using computer and data science tools such as README files for the first time. We believe there is a rich design space to explore how such tools can be expanded to scaffold novice learning in and out of schools. Finally, this work highlights how data cleaning, a process typically relegated to the background of discussions in data science education, can be an empowering activity that merits further research.

Acknowledgement

We would like to thank all the workers at DataWorks who entrusted us with their lived experiences and who generously shared their insights and time with us in ways that have resulted in deep friendships and a shared commitment to bettering data in our communities. We would also like to thank each of the reviewers of this paper who provided critically important feedback at all stages in the review process. This work was supported by the National Science Foundation under Grant No. 1951818. Any opinions, findings, and conclusions or recommendations expressed in this material are our own and do not necessarily reflect the views of the National Science Foundation.

References

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247-286. https://doi.org/10.1080/10508406.2011.611446

Aragon, C., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-centered data science*. MIT Press.

Bakker, A., Hoyles, C., Kent, P., & Noss, R. (2006). Improving work processes by making the invisible visible. *Journal of Education and Work, 19*, 343–361. https://doi.org/10.1080/13639080600867133

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). Beyond data literacy: reinventing community engagement and empowerment in the age of data. *Data-Pop Alliance White Paper Series*. Harvard Humanitarian Initiative, MIT Lad and Overseas Development Institute Internews. http://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf

Brown, Q., Grandison, T., Jenkins, O. C., Burge, J. D., Dillahunt, T., Thomas, J. O., Erete, S., & Rankin, Y. A. (2021). Reflecting on being black in computing. *Interactions, 28*(5), 34-37. https://doi.org/10.1145/3480237

Crooks, R., & Currie, M. (2021). Numbers will not save us: Agonistic data practices. *The Information Society*, *37*(4), 201-213. https://doi.org/10.1080/01972243.2021.1920081

Crutchfield, R. D., Skinner, M. L., Haggerty, K. P., McGlynn, A., & Catalano, R. F. (2012). Racial disparity in police contacts. *Race Justice*, *2*, 1-21. https://doi.org/10.1177/2153368712448063

Card, S. K., Mackinlay, J. D., & Schneiderman, B. (1999). *Readings in information visualization: using vision to think.* Morgan Kaufmann.

Charmaz, K. (2006). *Constructing grounded theory: A Practical guide through qualitative analysis*. Sage.

Computing Research Association. (2019). *CRA Taulbee survey*. https://cra.org/resources/taulbee-survey/

Davis, N. R., Vossoughi, S., & Smith, J. F. (2020). Learning from below: A Micro-ethnographic account of children's self-determination as sociopolitical and intellectual action. *Learning, Culture and Social Interaction, 24*. https://doi.org/10.1016/j.lcsi.2019.100373

DesPortes, K., McDermott, K., Bergner, Y., & Payne, W. (2022). "Go [ing] hard... as a woman of color": A case study examining identity work within a performative dance and computing learning environment. *ACM Transactions on Computing Education*, *22*(4), 1-29. https://doi.org/10.1145/3531000

D'Ignazio, C., &. Klein, L. (2020). *Data feminism.* The MIT Press.

Dörk, M., Feng, P., Collins, C., & Carpendale, S. (2013). Critical InfoVis: Exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)* (pp. 2189–2198). https://doi.org/10.1145/2468356.2468739

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proceedings of the National Academy of Sciences*, *116*(34), 16793-16798. https://doi.org/10.1073/pnas.1821204116

Eglash, R., Bennett, A., Cooke, L., Babbitt, W., & Lachney, M. (2021). Counter-hegemonic computing: Toward computer science education for value generation and emancipation. In *ACM Transactions on Computing Education (TOCE), 21*(4), 1-30. https://doi.org/10.1145/3449024

Epp, E. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled over: How police stops define race and citizenship*. University of Chicago Press.

Fox, S. E., Khovanskaya, V., Crivellaro, C., Salehi, N., Dombrowski, L., Kulkarni, C., Irani, L., & Forlizzi, J. (2020). Worker-centered design: Expanding HCI methods for supporting labor. In *Proceedings of the 2020 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1-8). https://doi.org/10.1145/3334480.3375157

Freire, P. (1968). *Pedagogy of the oppressed*. Routledge.

Glaser, B. G., & Straus, A. L. (1967). *The Discovery of grounded theory: Strategies for qualitative research*. Aldine.

Gray, C. M., Obi, I., Chivukula, S. S., Li, Z., Carlock, T., Will, M., Pivonka, A. C., Johns, J., Rigsbee, B., Menon, A. R., & Bharadwaj, A. (2022). *Practitioner trajectories of engagement with ethics-focused method creation*. https://doi.org/10.48550/arXiv.2210.03002

hooks, b. (1988). *Talking back: Thinking feminist, thinking black*. Between the Lines.

Jayathirtha, G., Shaw, M. Kafai, Y. B., & Lui, D. (2020). When a glove becomes a gun: From personally meaningful to socially critical restorying in maker activities. In *Proceedings of the FabLearn 9th Annual Conference on Maker Education* (pp. 94-97). https://doi.org/10.1145/3386201.3386205

Johnson, B., Shapiro, B. R., DiSalvo, B., Rothschild, A., & DiSalvo, C. (2021). Exploring Approaches to Data Literacy through a Critical Race Theory Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI'21) (pp. 1-15). ACM. https://doi.org/10.1145/3411764.3445141

Kahn, J. (2020). Learning at the intersection of self and society: The Family geobiography as a context for data science education. *Journal of the Learning Sciences, 29*(1), 57–80. https://doi.org/10.1080/10508406.2019.1693377

Lorde, A. (2007). *Sister outsider essays and speeches by Audre Lorde*. Ten Speed Press.

Lupi, G. (2017). Data humanism: The Revolutionary future of data visualization. *Print Magazine, 30*(3).

Matuk, C., DesPortes, K., Amato, A., Vacca, R., Silander, M., Woods, P. J., & Tes, M. (2022). Tensions and synergies in arts-integrated data literacy instruction: Reflections on four classroom implementations. *British Journal of Educational Technology, 53*, 1159-1178. https://doi.org/10.1111/bjet.13257

Matuk, C., Yoon, S., Polman, J., Amato, A., Barton, J., Bulalacao, N. M., Cafaro, F., Haldar, L. C., Cottone, A., Cortes, K., DesPortes, K., Erickson, T., Finzer, W., Taylor, K. H., Herbel-Eisenmann, B., Graville, C., Gutiérrez, K., Higgins, T., Himes, B., Lanouette, K., Lee, H., Lim, V., Lopez, M. L., Lyons, L., Milz, D., Olivares, M. C., Osche, E., Parikh, T. S., Philip, T., Rubel, L., Shelley, J., Rivero, E., Roberts, J., Roberto, C., Petrosino, T., Rubin, A., Shim, J., Silander, M., Sommer, S., Stokes, D., Tes, M., Trajkova, M., Urbanowicz, R., Vacca, R., Van Wart, S., Vasudevan, V., Wilkerson, M., & Woods, P. J. (2020). Data literacy for social justice. In *Proceedings of the 14th International Conference of the Learning Sciences* (Vol. 1, pp. 343-349). International Society of the Learning Sciences.

McNeil, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., & Erickson, T. (2019). How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). https://doi.org/10.1145/3290605.3300356

Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. D. (2006). Learning as a cultural process: Achieving equity through diversity. In R. K. Sawyer (Ed.), *The Cambridge Handbook of: The Learning Sciences* (pp. 489-504). Cambridge University Press.

Noss, R., Hoyles, C., & Pozzi, S. (2000). Working knowledge: Mathematics in use. In A. Bessot, & J. Ridgway (Eds.), *Education for mathematics in the workplace* (pp. 17-35). Kluwer Academic.

Pargman, C. T., McGrath, C., Viberg, O., Kitto, K., Knight, S., & Ferguson, R. (2021). Responsible learning analytics: creating just, ethical, and caring. In *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21)*. The Open University.

Passi, S., & Jackson, S. (2018). Trust in data science: Collaboration, translation, and accountability in corporate data science projects. In *Proceedings of the ACM on Human-Computer Interaction* (Article 136, pp. 1-28). https://doi.org/10.1145/3274405

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction, 34*(4), 361-388. https://doi.org/10.1080/07370008.2016.1210418

Rankin, Y. A., & Henderson, K. K. (2021). Resisting racism in tech design: Centering the experiences of black youth. In *Proceedings of the ACM on Human-Computer Interaction* 5, *CSCW1* (pp. 1-32). https://doi.org/10.1145/3449291

Rogoff, B. (2003). *The Cultural nature of human development*. Oxford University Press.

Rothschild, A., Meng, A., DiSalvo, C., Johnson, B., Shapiro, B. R., & DiSalvo, B. (2022). Interrogating data work as a community of practice. In *Proceedings of the ACM on Human-Computer Interaction*, *CSCW2* (Vol. 6, Article 307). https://doi.org/10.1145/3555198

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making space for place: Mapping tools and practices to teach for spatial justice. *Journal of the Learning Sciences*, *26*(4), 643-687. https://doi.org/10.1080/10508406.2017.1336440

Shapiro, B. R., & Pearman II, F. A. (2017). Using the interaction geography slicer to visualize New York City stop & frisk. In *Proceedings of the IEEE VIS 2017 Arts Program, VISAP'17* (pp. 1-8). https://doi.org/10.1109/VISAP.2017.8282370

Shapiro, B. R., Meng, A., O'Donnell, C., Lou, C., Zhao, E., Dankwa., B., & Hostetler, A. (2020). Re-shape: A Method to teach data ethics for data science education. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI'20) (pp. 1-13). ACM. https://doi.org/10.1145/3313831.3376251

Silvis, D., Clarke-Midura, J., Shumway, J. F., Lee, V. R., & Mullen, S. (2022). Children caring for robots: Expanding computational thinking frameworks to include a technological ethic of care. *International Journal of Child-Computer Interaction, 33*, 100491. https://doi.org/10.1016/j.ijcci.2022.100491

Smith Lee, J. R., & Robinson, M. A. (2019). "That's my number one fear in life. It's the police": Examining young black men's exposures to trauma and loss resulting from police violence and police killings. *Journal of Black Psychology*, *45*(3), 143-184. https://doi.org/10.1177/0095798419865152

Solomon, A. S. (2021). *Embodiment in computer science learning: How space, metaphor, gesture, and sketching support student learning* (Unpublished doctoral dissertation). Georgia Institute of Technology, Atlanta, GA. http://hdl.handle.net/1853/64777

Star, S. L., & Strauss, A. L. (1999). Layers of silence, arenas of voice: The Ecology of visible and invisible work. *Computer-Supported Cooperative Work (CSCW): The Journal of Collaborative Computing, 8*, 9-30. https://doi.org/10.1023/A:1008651105359

Suchman, L. (1995) Making work visible. *Communications of the ACM*, *38*(9), 56-64. https://doi.org/10.1145/223248.223263

Taylor, K. H., & Hall, R. (2013). Counter-mapping the neighborhood on bicycles: Mobilizing youth to reimagine the city. *Technology*. *Knowledge and Learning, 18*(1-2), 65-93. https://doi.org/10.1007/s10758-013-9201-5

Tygel, A.F., Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics*, *12*(3), 108-121. https://doi.org/10.15353/joci.v12i3.3279

Vakil, S. (2018). Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard Educational Review*, *88*(1), 26-52. https://doi.org/10.17763/1943-5045-88.1.26

Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and counterscripts in community-based data science: Participatory digital mapping and the pursuit of a third space. *Journal of the Learning Sciences*, *29*(1), 127-153. https://doi.org/10.1080/10508406.2019.1693378

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The Logic of everyday sense-making. *Journal of Research in Science Teaching, 38,* 529-552. https://doi.org/10.1002/tea.1017

Wilkerson, M., & Polman, J. (2020) Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences, 29*(1), 1-10. https://doi.org/10.1080/10508406.2019.1705664

Wolff, A., Gooch, D., Cavero Montaner, J. J, Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, *12*(3), 9-26. https://doi.org/10.15353/joci.v12i3.3275

Yadav, A., Heath, M., & Hu, A. D. (2022). Toward justice in computer science through community, criticality and citizenship. *Communications of the ACM*, *65*(5), 42-44. https://doi.org/10.1145/3527203

Yin, R. (2009). *Case study research. Design and methods* (4th ed.). Sage.

Arastoopour Irgens, G., Adisa, I., Bailey, C., & Vega Quesada, H. (2022). Designing with and for Youth: A Participatory Design Research Approach for Critical Machine Learning Education. *Educational Technology & Society*, *25* (4), 126-141.

Designing with and for Youth: A Participatory Design Research Approach for Critical Machine Learning Education

Golnaz Arastoopour Irgens*, Ibrahim Adisa, Cinamon Bailey and Hazel Vega Quesada
Clemson University, U.S.A // garasto@clemson.edu // iadisa@clemson.edu // cinamob@clemson.edu // hvegaqu@clemson.edu
*Corresponding author

ABSTRACT: As big data algorithm usage becomes more ubiquitous, it will become critical for all young people, particularly those from historically marginalized populations, to have a deep understanding of data science that empowers them to enact change in their local communities and globally. In this study, we explore the concept of critical machine learning: integrating machine learning knowledge content with social, ethical, and political effects of algorithms. We modified an intergenerational participatory design approach known as cooperative inquiry to co-design a critical machine learning educational program with and for youth ages 9 - 13 in two after-school centers in the southern United States. Analyzing data from cognitive interviews, observations, and learner artifacts, we describe the roles of children and researchers as meta-design partners. Our findings suggest that cooperative inquiry and meta-design are suitable frameworks for designing critical machine learning educational environments that reflect children's interests and values. This approach may increase youth engagement around the social, ethical, and political implications of large-scale machine learning algorithm deployment.

Keywords: Critical data science, Machine learning, Algorithmic bias, Participatory design research, Community youth program

1. Introduction

The world is becoming increasingly reliant on digital technologies to navigate our lives. Such technologies often collect, store, and analyze data to improve efficiency and quality of life. Data is increasingly being utilized in the development of machine learning (ML) and artificial intelligence (AI) systems. These systems are often used to make critical decisions about people and communities. For example, ML algorithms are now used to predict cancer patient outcomes, measure natural disaster damage using social media postings, and quantify traffic dynamics to reduce air pollution (Data Science for Social Good, 2021). However, such reliance on ML algorithms can also be problematic. For example, widely used internet search algorithms that distribute knowledge to billions of people daily have been shown to reinforce racial and gender discrimination at large scales (Noble, 2018), and government software has been used to extrapolate data that preemptively criminalize the actions of low-income populations (Eubanks, 2018). In addition to having harmful effects on marginalized populations, the design and operation of ML algorithms are invisible to the public (O'Neil, 2016), and many young people are unaware of how such algorithms collect and use their personal data (Pangrazio & Selwyn, 2019). Thus, it is essential to engage youth early on in their education about ML concepts and how to think critically about the social, ethical, and political issues around modern large-scale algorithm deployment that can (re)enforce inequities and harm marginalized populations.

In this study, we explore how youth ages 9 - 13 and researchers co-designed a critical machine learning program implemented in their after-school program, discussing the triumphs and tensions that emerged from this process.

2. Background and theory

2.1. Critical machine learning education

Artificial intelligence is a broad field about how computers use data, symbolic rules, and numeric models to analyze their environment and behave in ways that generally require human intelligence (Boucher, 2020). One of the ways humans implement AI is through machine learning (ML), which employs algorithms and processes to allow computers to solve problems and make intelligent human decisions. Humans build ML models by training computers to understand and recognize patterns from large datasets. These models are, in turn, used to make predictions and automated decisions about people and systems (Samoili et al., 2021). ML applications rely heavily on data and are employed to make automated recommendations and decisions that, in some cases, can be harmful to those from non-dominant populations such as women, people of color, and those living in poverty.

For example, Buolamwini and Gebru (2018) evaluated three commercial image classification systems used for facial recognition technology. The study was spurred by Buolamwini's personal experiences of being misidentified when using facial recognition software.

The researchers found that darker-skinned females were the most misclassified group, with error rates up to 34%, while the maximum error rate for lighter-skinned males was 0.8%. These error rates become particularly concerning when facial recognition systems are being used by U.S. law enforcement and other government agencies to detect unlawful behaviors. Buolamwini and Gebru's (2018) study suggests that those who reside on the margins of society continue to be marginalized with biased datasets used for training ML algorithms. Thus, for youth to assess such technologies and advocate for themselves when models make unfair decisions against them, they need to understand the basics around how ML works and be able to reflect upon the risks and consequences. In our prior work, we have proposed a critical machine learning (CML) educational approach (Arastoopour Irgens et al., 2022) that integrates vital pedagogy (Freire, 1970; Giroux, 1985) into computer science and machine learning education. In this integrated approach, learners ask questions such as, who develops these technologies? What types of data are used to train machines? What is the history behind the data used? What decisions are made based on the outputs of the algorithms? These questions facilitate reflections, discussions, and actions around disrupting oppressive paradigms related to the deployment of ML-based technologies.

In recent years, researchers have explored approaches for engaging youth in learning about the social, ethical, and political implications of ML and AI. Williams and colleagues (2019) discovered that young learners' understanding of AI concepts was based on the degree to which they were able to actively participate in hands-on activities such as modifying a robot's input, teaching a robot how to play a game, or training it to tell the difference between objects. Studies with middle-school-aged children have shown their abilities to engage with ethical AI concepts and identify the societal impacts of racist and biased algorithms (Ali et al., 2019). Classroom intervention research suggests that upper elementary school-aged students can consider oppression from multiple perspectives, including the broader historical framework of how society is organized and how to create change (Fain, 2008), analyze and interrogate literature around societal issues such as immigration (Braden, 2019), and address and challenge social inequities in their curricula (Kersten, 2006). To prevent children from internalizing harmful messages about oppression, it is essential to proactively engage them in reflections early (Boutte & Muller, 2018), especially those who belong to groups that experience discrimination (Ayón, 2016).

Although there is evidence that middle school-aged children are capable of reflecting on how AI technologies can be unfair to themselves and others, discomfort and tensions may arise when discussing issues of AI oppression with youth. For example, Lee and colleagues (2021) worked with community partners and framed their activities with sensitivity when working with students from populations that are underrepresented in STEM and computing. Because many of the activities directly addressed discrimination against students from these groups, the researchers framed the discussion on how the AI community can solve such problems rather than pressuring the students themselves to solve AI bias problems. The researchers also noted a need to include more positive examples of AI technologies in healthcare, education, and art rather than exclusively discussing negative implications.

This empirical work demonstrates initial progress on critical ML education. It suggests that learners are likely to develop essential skills and an understanding of ML when engaged in hands-on activities that are personal, agentic, and sensitive to their identities. However, there is still a lack of consensus on how to design environments to support such learning (Wolff et al., 2019) that is meaningful for learners and their communities.

2.2. Participatory design research

To engage youth in *culturally sustaining pedagogies* (Paris, 2012; Paris, 2021), in which we maintain the culture of their communities while simultaneously providing access to dominant ways of thinking around critical ML, it is necessary to include youth in the design of their learning. Traditionally, children do not play a significant role in the creation of their learning experiences. However, not including children's input in the design of their educational experiences leads to the possibility of a mismatch between designers' intentions and learners' interpretations. When youth are both users and designers, they engage in mutual learning and co-construction of knowledge with instructors, mentors, and other stakeholders (Robertson & Simonsen, 2013). Moreover, incorporating children's cultural values may increase the chances of sustained engagement and learning compared to only designing for fleeting youth interests (DiSalvo & DesPortes, 2017).

Druin (1999) argues that designing with and for children requires reimagining how to facilitate design processes to include children's voices effectively. Based on empirical work, Druin proposes *Cooperative Inquiry*, comprising three techniques: contextual inquiry, technology immersion, and participatory design. In contextual inquiry, adults and children collect data about how children and adults interact in a selected environment. These notes and data inform the creation of the technologies and programs. During technology immersion, children "tinker" with multiple, novel technologies in their own environment. Such technology-rich, time-intensive experiences allow researchers to observe multiple patterns of children's activity. The participatory design component involves adults and children creating low-tech prototypes of designs using materials such as sticky notes, clay, string, paper, or markers. One central assumption in participatory design and cooperative inquiry is that there are multiple and valued forms of expertise stakeholders bring to the design process. For example, in intergenerational design, children are experts in what it means to be a child today (Guha et al., 2013). Children can use their imaginations to propose creative design ideas that may inspire adults. Many of the child's ideas cannot be realized in the actual design process, but adult partners can help reformulate the ideas so that they are workable with existing technologies. Interactions of this form welcome children's ideas and give them agency in the design process.

However, one challenge that adult designers face is that no matter how much preparation goes into the design of open-ended technology-based learning activities, it is unclear how children will appropriate the technology. As Ehn (2008) puts it, a researcher's "envisioned use is hardly the same as actual use, no matter how much participation there has been in the design process" (p. 95). One approach to address this challenge is through meta-design, in which the tool is designed before users engage with it. Still, the design allows for flexibility such that users can act as co-designers and customize, extend, or redesign aspects of the tool (Fischer, 2021). Meta-design embraces a co-adaptive process between users and a system and provides opportunities, tools, and social reward structures to refine systems to fit users' needs. Through the lens of meta-design, adults and children do not necessarily design concurrently. Adults create a tool or learning activity that offers opportunities for creative production and modification of tools and procedures. In a learning context where adult designers are also educators, both educator and child learner play the role of meta-designer. This means all participants fluctuate between the roles of learners, designers, and contributors (Fischer et al., 2004). The activities are not finished products, and learners are informed participants who have the power to shift the learning goals and methods. The meta-design approach aligns well with interactive networked digital technologies, such as Scratch (https://scratch.mit.edu/) and ML-specific authoring tools, such as Google's Teachable Machine, that allow children to be consumers and producers of media (Jenkins, 2006). In particular, the sensitive and reflective nature of exploring ML tools through a critical lens with children also lends itself to the reflective, asynchronous design and implementation that occurs in meta-design. However, few studies have explored the roles of children and adults as meta-designers as they engage with ML tools through a critical lens, informing the design of learning environments.

In a previous study, we suggested that the youth who participated in this co-designed CML program made more sophisticated connections with socio-political orientations and ML content as they progressed through the program. They engaged in computational practices, such as experimenting and iterating, testing and debugging, reusing and remixing, and abstracting and modularizing (Arastoopour Irgens et al., 2022). In this current study, we focus on the program design that facilitated such learning. We rely on cooperative inquiry techniques and meta-design approaches to explore how children and adults partnered as meta-designers in a CML educational program co-design.

The research question in this study is: *How did children and adults engage as meta-design partners during the CML program?*

3. Methods

3.1. Context, participants, and researcher positionalities

We implemented the project in after-school programs at two centers: Green Community Center and Sunshine Community Center (the names of the centers and children are pseudonyms). Both centers serve elementary schools in a Southern U.S. County that has a mix of urban and rural areas, a poverty rate of 13.4%, a household median income of $56,609, and a population that is 67% White (non-Hispanic), 23% Black, 5% Hispanic, and 2% Asian. Participants included 44 youth ages 9 – 13, 3 staff counselors, and 4 researchers. Each youth participant used their school-assigned Chromebook. The youth population consisted of Black, Latino, and White children, with a mix of those who presented as girls and boys. Youth attendance was variable. The researchers were university faculty and graduate students and consisted of a White/Middle Eastern woman, a White woman from the local region, a Nigerian Black man, and Costa Rican Latina woman. All researchers are actively opposed to big-data algorithms that convey and perpetuate historical and current racism and sexism at large scales. The lead author and director of the research project is a former computer science and mathematics instructor whose perspective has influenced the design of the current program.

3.2. Design of the CML educational program

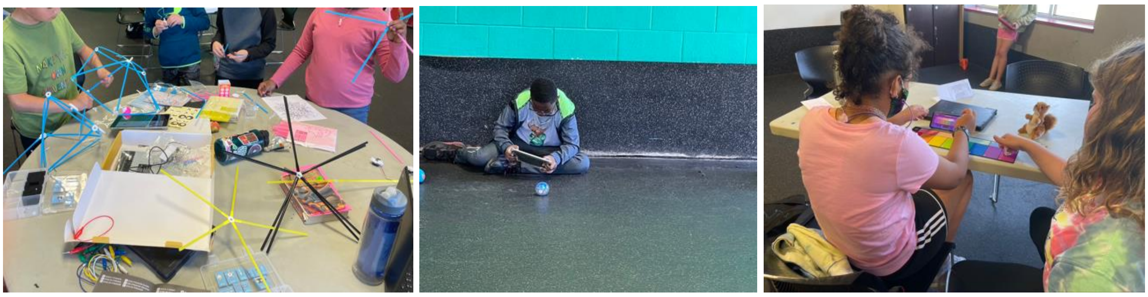
The program spanned 15 days, was implemented 2-3 days a week and occurred in three phases: *Initial Exploration, Activities and Discovery,* and *Youth Design of Machines.*

3.2.1. Phase I: Initial exploration

During this phase, the objectives included: deciding where and how to implement the program initially; building relationships with youth and staff; determining youth values and interests; analyzing the program and adjusting as needed; and determining youth baseline knowledge about algorithms and ML.

Before any CML activities were fully designed and implemented, researchers volunteered at the after-school program. We planned to spend four days in this role observing children and staff in their day-to-day activities at the center, assisting children with their homework, and engaging in casual conversation to build relationships and get to know one another. According to our observations, the typical schedule of activities at the centers was staggered arrivals from different schools (10 minutes), homework time (45 minutes), and playing outside (until a parent arrived). At Green Community Center, two staff members sat at the front of the room watching children and occasionally shouting if they violated the rule of having three at a table. This rule was implemented because of the COVID-19 pandemic restrictions in the spring of 2021. Because of the global health crisis at the time, all students and adults wore masks during the implementations, and the after-school center shut down for one week in the middle of our schedule after a COVID-19 outbreak.

Figure 1. (a) children creating figures using Strawbees straws and connectors with optional robotics, (b) a child programming and moving a Sphero robot ball using a tablet, (c) children using the Specdrums app and physical kit to create music remixes



(a) (b) (c)

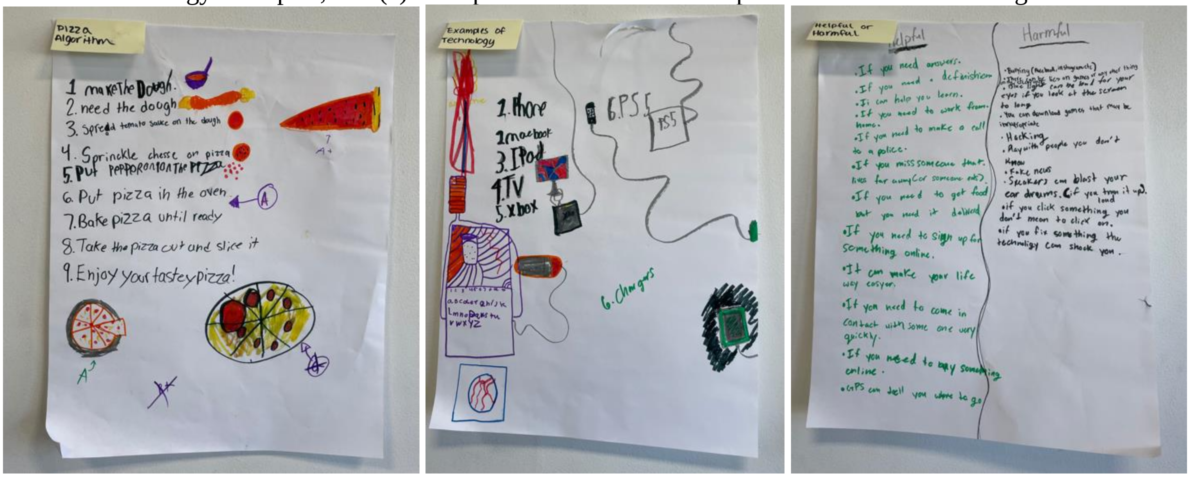
After we learned that most of the children did not have homework, we brought robotics toys during the next three volunteering as an alternative way to engage with the children and observe them with technologies. All children engaged with at least one of the toys. They built robotic sculptures with the Strawbees straws (https://strawbees.com/), created musical remixes using the Specdrums kit (https://sphero.com/collections/all/family_specdrums), and organized racing and bowling competitions with the Sphero (https://sphero.com/) robot balls (Figure 1).

3.2.2. Phase II: Activities and discovery

During the second phase, the objectives included: continuing to grow relationships with youth and staff; continuing to discover youth and staff values and interests and implementing them into the activities; developing and supporting mutual learning through group activities and discussions; providing activities and tools that could assist youth in knowledge construction and flexible design.

We created and adapted activities from previous studies (Bailey et al., 2021), MIT's How to Train Your Robot Curriculum (https://httyr.media.mit.edu/), and MIT's AI Ethics Education Curriculum (https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/). Each day, the activities and discussions systematically built on the youths' prior experiences and the knowledge they had gained. We started the implementation by explaining to the youth that algorithms are instructions. They were then asked to make a pizza algorithm in teams using markers and giant sticky notes. By listing directions for making a pizza, they visualized algorithms as instructions and questioned others' pizza algorithms (Figure 2a). In the same teams, the youth answered two questions: (1) What are some examples of technology you use or see throughout the day? (Figure 2b) and (2) What are some ways we use these helpful and harmful technologies? (Figure 2c). After creating their posters, the youth walked around the space to view other teams' posters, wrote comments on sticky notes, and attached them to the posters.

Figure 2. (a) Example of one team's pizza algorithm visualization, (b) example of another team's everyday technology examples, and (c) example of another team's helpful and harmful technologies list



(a) (b) (c)

Next, the youth used one of Google's AI experiments: Quick, Draw! (https://quickdraw.withgoogle.com/). In this application, users are asked to draw an object, such as a guitar or rainbow, and the algorithm guesses the object. The algorithm was trained using a neural network and a training dataset with millions of drawings from global users. After experimenting with the Google Quick, Draw!, youth engaged with researchers in discussing the potential bias embedded in the tool.

In the next activity, the youth used Google's Teachable Machine (https://teachablemachine.withgoogle.com/) to classify images of cats versus dogs. Youth were told they were creating "a facial recognition software to determine whether a pet is a cat or dog" and given two envelopes containing printed images of cats and dogs. One envelope was labeled "training data," and the other was labeled "test data." Using the training dataset, youth used their Chromebook webcam to train the Teachable Machine to differentiate images of cats and dogs. However, the training dataset contained a more extensive variety and higher quantity of cats; thus, the resulting trained machine misclassified dogs more often than cats. Through this experience, youth engaged directly with training data, test data, and bias in training sets.

3.2.3. Phase III: Youth design of machines

During this next phase, objectives included: providing youth with the tools they needed to design their own teachable machine; guiding youth in understanding algorithm bias; supporting youth as they designed and built (1) their own teachable machines and (2) trained their own robots; and supporting youth choice in their two designed products.

In this phase, the youth first designed a machine using Google Teachable Machine individually or in teams and chose their input data to train their machine. They presented their machines to their peers, staff, and directors and had the opportunity to earn one of two prizes for "most creative machine" and "most functional machine."

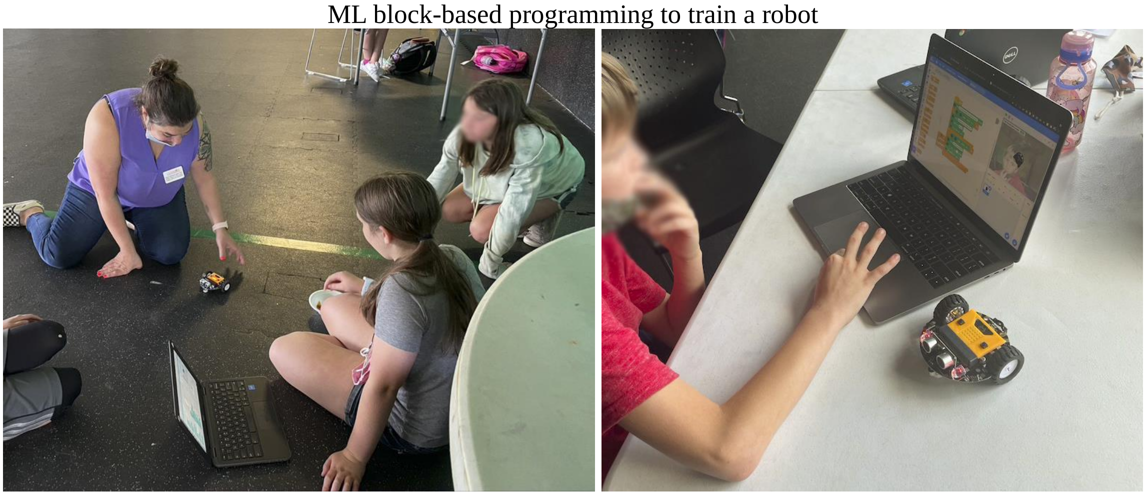
Figure 3. Summary of CML education program activities and meta-design processes (*Note.* Numbers represent the days. Red Circles = Phase I: Initial Exploration, Blue Circles = Phase II: Activities and Discovery, Green Circles = Phase III: Youth Design of Machines.)

PHASE	DAYS	CML EDUCATION PROGRAM ACTIVITIES
Initial Exploration	1 - 5	Volunteer • Researchers and youth build relationships • Youth experiment and play with digital technologies in their own space, in their own ways.
	6	Pre-Drawing & Interest Artifact Activity • Youth express their knowledge about algorithms and ML through drawings • Youth express interests and values through creating artifacts and discussing
Activities and Discovery	7	Google Teachable Machine Self-Directed Activities • Youth engage with our designed website containing Google Teachable Machine tutorials and written directions to complete their own machines.
		Major Re-Design
	8	Pizza Algorithm & Harmful/Helpful Tech Activities • Youth are introduced to algorithms with definitions and examples • Youth use guided reflections to draw connections between algorithms and their everyday experiences
	9	Google Search & Google QuickDraw Activities • Youth are introduced to machine learning with examples • Youth use guided reflections to define ML algorithms and how training machines in particular ways results in inequitable outcomes
	10	Google Teachable Machine Redux • Youth watch a demonstration of a simple Teachable Machine • Youth unknowingly created a biased Teachable Machine • Youth use guided reflections to understand why their machine was biased and the consequences of biased algorithms • **META-DESIGN:** Co-develop the upcoming Teachable Machine competition
Youth Design of Machines	11	Teachable Machine Competition • Youth design and build a Teachable Machine aligned with their interests and values • Youth receive prizes for different categories with community judges • **META-DESIGN:** Co-develop the machines to integrate youth interests
	12	Coded Bias Video & Superhero Robot Stories • After the viewing of the Coded Bias trailer, youth engage in guided discussion about algorithm bias in facial recognition: who designs and who gets harmed • Youth reimagine ML technology for social good that mitigates bias. • **META-DESIGN:** Co-reflect on how to publicize youth work
	13 - 14	ML Robot Competition • Youth create a prototype of their Superhero robot with Scratch, Teachable Machine, and YahBoom robots • Youth receive prizes for different categories with community judges • **META-DESIGN:** Co-develop ML robots and collectively discuss social justice goals
	15	Post-Drawing & Cognitive Interviews • Youth express their knowledge about algorithms and ML through drawings • **META-DESIGN:** Co-reflect on the design of the set of activities

Before their next designed product, youth watched the *Coded Bias* film trailer (Kantayya, 2020), which featured Joy Buolamwini's realization of racist facial recognition technologies. Afterward, youth, staff, and researchers discussed racial discrimination embedded in ML technologies. After thinking critically about harmful technologies, youth shifted perspectives and were asked to create imaginative stories about helpful robots.

For their final design, the youth experimented with Yahboom! Micro:bit Robots using block-based programming software from MIT's How to Train your Robot Curriculum. They were asked to use ML algorithms to train a robot to be helpful to society (Figure 4). The ML concepts embedded in the block-based language included incorporating a Google Teachable Machine and voice/text classification. Before designing and programming their robots, we provided them with a demonstration of a trained image classifier using the laptop's webcam that detected whether a human face was present or not. If the webcam detected a human face, the robot would print a smiley face on its display, shine magenta headlights, and spin around in circles. If the webcam did not detect a human face, the robot would print a sad face on its display, turn off the headlights, and move backward. Youth presented their robots to their peers, staff, and directors of the program and could earn prizes. Figure 3 summarizes the CML education activities and co-design processes by day.

Figure 4. (a) One researcher and two children working together to train and test a robot and (b) one child using ML block-based programming to train a robot



(a) (b)

3.3. Data collection and analysis

During the implementation sessions, at least one researcher in the session kept a reflective research journal and took observational field notes. As Glesne (2014) suggested, the reflective research journal documented the researcher's reflections on tensions, successes, and other feelings as we interacted with the children and staff. In parallel, the observational notes documented the researcher's evidence from observations and corresponding analytical notes. The researchers interviewed participants who submitted signed consent forms from their parents. We followed a mixed cognitive clinical (Russ et al., 2012) and semi-structured interview protocol (Appendix A). In the first section of the interview, the researcher displayed the learner's teachable machines and asked learners to reflect on their work. In the last half of the interview, the children were asked about their thoughts and feelings toward their experiences in the program and what they would add or change in future iterations. We collected artifacts from learners, including their completed teachable machines, superhero robot block-based code, and photos of giant sticky note activities (pizza algorithm, types of technologies, harmful/helpful technologies, superhero robot stories). We also took pictures and videos during their work time.

These data were analyzed using a descriptive phenomenology methodology (Creswell & Poth, 2016), which describes the lived experiences of a group of people. We used thematic analysis (Braun & Clarke, 2006) to describe the meta-design phenomena in this specific CML education context. Specifically, the analytic phases included familiarizing with the data, generating codes, constructing themes, reviewing/reflecting on themes, defining themes, and producing the report (Terry et al., 2017). The findings describe what the children and adults experienced when engaged in the CML program and how they experienced it through the lens of meta-design.

4. Results

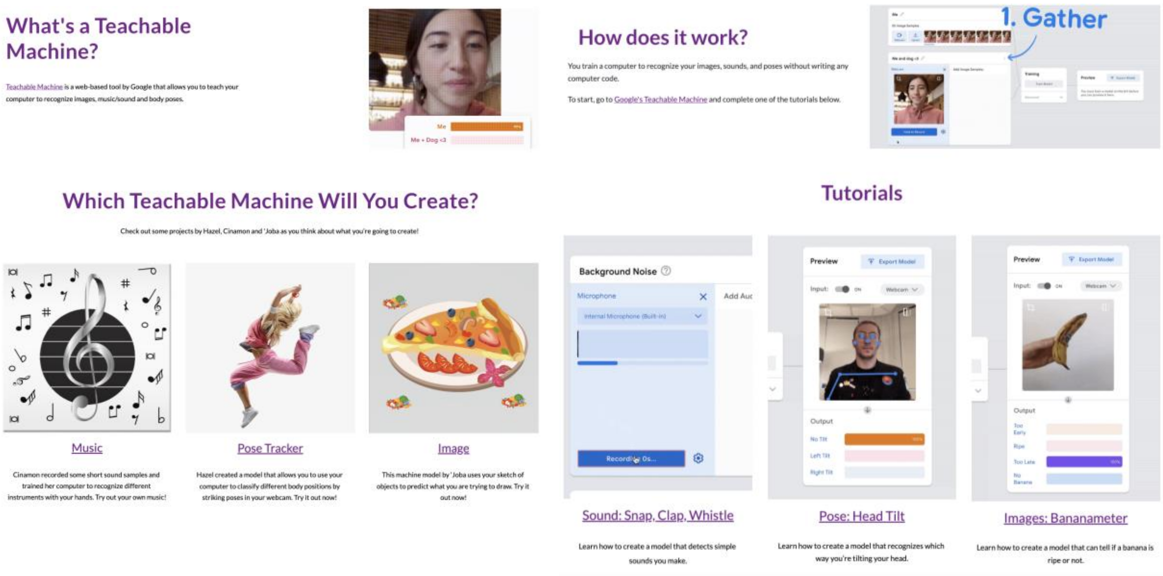
Through a meta-design conceptual lens, we analyzed the observational field notes, cognitive interviews, and youth artifacts. We identified 3 themes around adults and youth as meta-designers of a CML education program and what conditions facilitated co-design.

4.1. Researchers and youth persisting through failures and redesigning in real time

Following participatory design methods, we first volunteered at the after-school center to build relationships with the youth and understand how they interacted in their space. According to our observational notes, youth spent their time playing computer games, completing crossword puzzles, or conversing in small groups. When the weather was suitable, staff escorted the youth to play outside. On the second day at Green Community Center, the youth invited researchers to a fort they were building in a nearby wooded area. Youth included researchers in their imaginative play and used branches to continue building the fort.

Based on our observations and discussions with the youth, we designed self-directed and creative activities. We developed a website that linked to Google Teachable Machine, linked to tutorials, and linked to examples created by the research team (Figure 5). We added a discussion board to the site, which asked the youth to post links to their machines and discuss. On day 7 (see Figure 3), we implemented this activity at Green Community Center by taping printouts of a QR code on tables that led youth to the website. We planned to give youth access to the website and have them self-direct their experience, similar to what one would experience in other informal learning settings such as museums.

Figure 5. Two pages from our initial website for youth to explore Google Teachable Machine and design their own ML machines



Unfortunately, as one researcher put it afterward, “It was a complete failure.” According to researcher notes from day 7, “They [the youth] said they did not know anything about algorithms and machine learning… Once on the website, they just scrolled down without reading and didn’t know what to do. A few of them [8 out of 20 youth] got to try Teachable Machine.” We also observed that most youth who experimented with the Teachable Machine quickly abandoned the tool if they did not receive guidance, feedback, or intervention from an adult. However, one group of three girls sustained engagement with the device and used their toy figurines and webcam to create an image classification system that distinguished between their toys. They tested and modified their machine throughout the afternoon to improve its classification capabilities.

The day after this implementation, the researchers decided to hold an emergency design meeting to redesign tools and activities for the youth that provided less information at one time, less text, more discussion, and direct interaction with adults. Based on our observations and discussions with the group of three girls who successfully created a machine and seemed to enjoy the process, we decided to directly connect to and actively support students’ everyday interests with ML. During this significant re-design between days 7 and 8 (see Figure 3), researchers created an outline of activities that gradually introduced youth to algorithms, ML, training datasets, test datasets, classification systems, bias in datasets, and harmful technologies. We adapted activities from MIT's How to Train Your Robot curriculum to better fit with our population of children. For example, one researcher noted in their observations and interactions with youth that "these kids like to have objects and things they can touch and manipulate." As a result, we printed out pictures of cats and dogs for the biased Google Teachable Machine activity. This way, children could hold physical photos to their webcams to train their machines rather than download and upload digital images.

Researchers also scheduled reflection discussions with the youth at the start and end of the sessions. These reflection discussions functioned as a pedagogical tool where youth could reflect on their learning and co-construct understanding with others. These reflection discussions also served as a meta-design tool where youth could express their desires for changes or modifications to the learning activities. These changes in the program engaged youth more deeply in the CML content because they could see the benefits for them in terms of participation and feel that their voices were being heard. For example, during one discussion, a researcher asked if one of the children could summarize what they did the previous day.

Justin summarized the Teachable Machine activity in which he trained a facial recognition system for cats and dogs with biased training data. He explained, "We went on this website and programmed a computer to see if it knew if it was a cat or dog. And it would actually say cat because it had more cat pictures instead of dogs. And every time you got a cat, you see more pop up [referring to the increased percentage of confidence in the classifier]." The researcher added, "Okay, so it was better for cats than dogs because there were a lot more cats than dogs in the training set." Several children added how they tested the machine on their faces, and the machine confidently classified them as either dogs or cats. The group laughed together when these stories were shared. Joe insightfully added that he was likely identified as a cat because his face was in the webcam's line of sight while training his cat classifier. The researchers did not anticipate that youth would test their *own faces* when creating a facial recognition system for cats and dogs. However, this exploration of their faces led to joyful interactions with the ML tool and playful discoveries around human error when training datasets. This portion of a reflection discussion highlights how researchers and youth created understanding together, both around CML knowledge and the design of the activities, when reflecting on their open-ended learning activities.

4.2. Technical and social conditions to facilitate researcher and youth participation in design

Our observations inspired another critical part of the significant redesign during the first five days of volunteering. We observed that all children engaged with at least one of the robotic toys we brought to the centers. One researcher noted that on the second day of robotics play that "since they knew that we were coming with the toys, they almost jumped and barely allowed us to walk. They spent the whole time with the toys… They even wanted to continue playing when it was time to clean up." Based on these observations and interactions, the researchers changed direction, made physical computing a major part of the program, and offered opportunities for children to build ML robotic machines of their interests with guidance from adults. Specifically, we chose a Scratch programming interface with a Teachable Machine extension which was used to program robots that used a micro:bit processor. Youth were asked to create narrative stories about robots that can be helpful to people, which we called "superhero" robots. Although their stories were fantastical in nature, with researchers' help, youth could translate aspects of their stories into a workable, programmable robot that used ML algorithms.

The youth were aware of the technical flexibility of the program and tools. In his interview, Lucas said he had no prior experience with programming and had not heard anyone talk about ML before. Lucas added, "When you started telling us those things [referring to bias in training datasets], I was surprised. So, now that I know this, about technological machines, I really like them because it's like, it doesn't have to be like anything in the world. You can be creative on your own and you can actually make something new." Lucas was referring to his learning about biased datasets. He explained that he enjoyed the ability to create a machine and contribute something to the world that was not there before. In turn, Lucas acted as a meta-designer, designing his own learning experience and designing a new ML application, which could inspire and change future iterations of the program. He continued to explain the benefits of participating in this program and how creativity is essential for his future profession, adding, "Cuz the way you work or something, you have to be creative. I want to be successful in life." Similarly, Emma specifically enjoyed the creative aspect of designing and programming a superhero robot. She noted, "I really liked the robots. And I also really liked when we got to come up with our hypothetical robots." Emma's use of the phrase "got to come up with" indicates her perspective that she had the agency to design her robot story. She stated that she also enjoyed "the drawing part of it and, like, kind of creativeness." Like Lucas, Emma valued the creative and flexible nature of the design phase and the ability to implement her design into an ML robot. Emma felt that what she learned in the program was valuable to her future education: “I feel as we go into middle school, or high school, college, and so on, we’re always going to have to know something like this.”

In addition to flexible technical conditions for participation, involvement social conditions encouraged youth to make sense of CML and reflect on their participation in the program with researchers. For example, researchers held discussions with youth about how to showcase their creations. Youth and researchers created a list of options for presenting their ML machines and robots: post on a website and share publicly, have a “science fair” style presentation, invite friends and family, or hold a competition with prizes. Ultimately, the youth chose to hold a contest in which staff from the after-school center were the judges. These conversations were one example of creating comfortable social conditions for broad participation in the design of activities.

Although many conversations among youth and adults were benevolent, some discussions around discrimination brought discomfort and distrust for some youth. The most notable examples come from discussions following the Coded Bias viewing of how ML algorithms harm marginalized communities. At Sunshine Community Center, when the researcher asked what was interesting to the youth about the film trailer, Carter, a White girl, responded, “This is talking about… racism, so it said that, like, it can change whether people get property or have to pay the same prices for things as others.” Here, Carter noticed “racism” and inequities regarding how “people” obtain housing or purchase goods when ML algorithms are involved. Kendall, an African American girl, responded to Carter, “What I thought was interesting was the same reason, just because of the software, the person wasn’t recognized… they could get locked out of their house, or they could be denied for housing.” Sitting close to Kendall, Justin, an African American boy, nodded and affirmed her statement.

Carter then clarified who the “people” were being discriminated against, and she used an outdated term for African Americans, which Justin found offensive. This caused Justin to exclaim, “What?” and put his head down on the table. The researcher did not engage further with this language and continued the conversation. Kendall also ignored Justin and continued, “I thought what was interesting was the fact that the problem was so big that she [Joy Buolamwini] had to take it to court. The researcher asked, “Yes, who did she say it was mostly representing…” Justin popped his head up and shouted, “White Men!” The researcher responded, “That’s right, Justin.” Here, Justin rejoined the discussion by using his own racial vocabulary.

The researcher then asked a direct question about biased training datasets, “What do you think the training data looks like for the stuff that Joy was using?” Kendall answered and used the language that Justin introduced, “Umm mostly White men, because they didn’t have any other people in there to help them like, create their software besides White men.” Justin whispered, “Racist.” The researcher directly addressed Justin this time, “Right, yeah. I mean, it was biased towards white men because those are the people making a lot of the software.” Kendall then put the pieces together and said, “Well, bias is also like being racist.” The researcher extended Kendall’s comment, “Bias is a general term. Racist is a specific term for being discriminatory towards people who are a certain color or a certain race and that’s what’s happening here, right?” Justin, Kendall, and Carter nodded and affirmed. As the conversation continued towards defining bias, racism, and how ML algorithms can be programmed to be anti-Black, Justin fluctuated between engaging passionately in the discussion and disengaging.

There was a similar dynamic at Green Community Center, in which some youth engaged in the discussion and others, both African American and White, withdrew from the conversation. Afterward, we realized the youth were leading the direction of the discussion based on what they noticed, and the researcher mainly was following. This youth-led discussion facilitated an open system in which youth could shape what they wanted to discuss around algorithm bias and racism and how they wanted to continue with the activities. On the other hand, adults’ lack of structure or setting of ground rules and common language before the discussion led to some discomfort and insensitive language. The open nature of the discussion led to situations in which the researcher was unprepared. In turn, after the Coded Bias discussions, the researchers did not mention discrimination, racism, or other sociopolitical contexts around ML unless an individual child said it first and wanted to discuss it further. We let the youth’s level of experience and interests around institutional racism direct how we pursued the implementation of the remaining activities. Some children, such as Kendall, had experience discussing issues of racism and discrimination, were passionate about the issues, and could apply their knowledge to ML models. We provided additional content knowledge with these youth and helped them negotiate and build upon their prior experiences.

4.3. Negotiations and tensions between researcher and youth goals

According to our research notes, our goals were for youth to program robots for social good that incorporated some form of ML and to explain how training datasets were biased. However, we wanted to provide a flexible, modifiable environment such that learners could accomplish the CML goals in creative ways that aligned with their interests and values.

The youth took advantage of the flexible learning environment by designing their own experiences. For example, before the program began, Bianca, Ian, and Eric founded a gaming club called the Super Phenomenal Gamers (pseudonym). They explained to researchers that they formed this club based on their common interests in watching competitive gamers stream videos and wanting to explore esports as a career option. All three children were inconsistently involved in the CML activities; some days, they would participate, and other days they would play games on their mobile devices or update their YouTube channel. However, at the end of the program, they decided to develop an ML robot that relied on speech recognition. The purpose of their robot was to promote the Super Phenomenal Gamers YouTube channel. When the machine was turned on, the robot asked, "Are you subscribed to the Super Phenomenal Gamers?" If a person responded "yes," the robot changed its headlight color to green, spun around several times, and replied, "Thanks for subscribing." If a person responded "no," then the robot remained still, changed its headlight color to red, answered, "Subscribe for more great content," and displayed "Why not?" on its scrolling marquee.

Because of his interest in streaming videos, Eric used his mobile device to record the team working on their robot. He also recorded the group receiving feedback from one researcher and presenting their robot to the judges during the competition. The children posted these videos on the Super Phenomenal Gamers' YouTube channel, which became an unexpected source of data, detailing how the children worked together when adults were not interacting or recording them. The videos provided Eric and his team with a method for documenting their work and formulating a narrative from their point of view. For example, when a researcher, Bianca, and Ian were debugging their code and ignoring the camera, Eric turned the camera to the researcher and said, "Say hi to the channel! People who go to Clemson you might know her, so please subscribe." Most importantly, the children made an unprompted design choice to record and post videos of their work because it directly connected to their goals of being competitive gamers. The videos and robot design promoted their YouTube channel that contained their gaming videos. Bianca, Eric, and Ian's choices illustrated the flexible nature of the ML tools and activities, which allowed them to show researchers alternative ways for the tools to be used in conjunction with other popular media in ways that the children valued. Although the Super Phenomenal Gamers enthusiastically used ML algorithms to program a robot and made creative choices for marketing purposes, the *critical* lens was missing from their project. These children did not design a robot for broader social good, nor did they reflect on algorithm bias and how to mitigate bias in their design. In this case, the researchers compromised their goal of having youth design for social good.

In contrast, other children designed robots for social good and were able to incorporate their interests and values. For example, Kendall decided to integrate her work using Google's Teachable Machine with her superhero robot project. Using the webcam on her laptop, she trained her machine to classify objects by color. During the cognitive interview, she explained that she built this machine to help children learn their colors, specifically reflecting on her younger cousin's lack of resources for learning. She explained, "My cousin… when she was growing up, she didn't have the opportunity to, like sit down every day and like watch, like TV shows that teach her colors and stuff. And so, the only time she had stuff to learn is when I came down with my books and stuff and like taught her. And so, I thought to myself, that could be happening to multiple other kids all over America. And so, I thought, well, maybe I could make a machine that can help kids with that." In that moment, it also occurred to her that her machine could be adapted to assist those with vision impairments. She imagined recreating her machine as a phone app that could help blind children identify colors: "They could download the app on to their phones… And if they're like doing something, and they need to know what color it is, they turn on the camera, and it sends a link to their phone. And for each color has a different frequency. For orange or for red, it would be high pitch. But for black, it would be really low to be able to tell which one it is."

When asked how she trained her machine, Kendall said she uploaded images with various colors from Google and tested the machine on "sticky notes, an orange, my backpack, and outfit." However, she noticed that her machine did not accurately classify objects with different textures. She explained, "at first, I just put in solid colors, but then I realized when I put my mask up to the camera that there's multiple different textures and stuff of different colors. So, for each color, I inserted a different texture." During the interview, she demonstrated the updated functionality of the machine by holding her hand up to the webcam: "Yeah, my hands are really wrinkly. But if I put my hand into the camera, I'd be able to recognize that my hand is brown." Kendall said she also walked around the center to test her friends' machines. She pointed out that other youth created training datasets that were biased and, in turn, were not as functional as she had expected. She referred to one person who “only used a picture of himself” to classify faces and concluded that “his data was biased.” In contrast, Kendall said her machine worked well and noted that she won the most creative teachable machine award at her center. Kendall’s work is an example of youth’s ability to design machines for social good and understand bias in training datasets and how to retrain data to minimize bias, all while aligning with their interests and values.

5. Discussion

In this study, we integrated cooperative inquiry (Druin, 1999; Guha et al., 2013) techniques with meta-design (Fischer, 2021) to explore how researchers and children interact as meta-design partners in the context of CML education. The findings in this study described the conditions that allowed adults and youth to be meta-designers of a CML education program, as well as the tensions and negotiations that emerge from an intergenerational design process involving sociopolitical contexts.

In our initial role as volunteers, we built reciprocal trusting relationships by being curious and valuing children’s ideas, ingenuity, and practices in their after-school space. Being a volunteer was similar to the *least-adult role* in Cumbo and colleagues’ (2019) study in which research is situated in the children’s familiar play environment, child-led interactions with adults shape the activities, and adults are reflexive about their changing relationship with the children. This phase of the relational building laid the foundation for engaging as meta-designer partners with children.

When the adult’s role changed to meta-designer, we developed tools and activities that we thought would meet our goal of youth creating robots for social good that incorporated ML and explaining the consequences of biased training datasets. However, our initial design assumptions led to a failed implementation in which youth were not interested in creating ML machines or exploring biased datasets. After discussing further with the youth, researchers redesigned the implementation plan. Like Williams and colleagues (2019), we discovered that young learners preferred physical computing and relied on adult-guided, hands-on activities to develop their understanding of ML. Aligning with the meta-design concept around developing open systems in which users can use products but also design them as they use them (Fischer & Scharff, 2000), we implemented technical tools and pedagogical activities such that children could modify their artifacts and learning experiences.

Moreover, we adapted the cooperative inquiry techniques of observing children use technologies in their own space (Druin, 1999) and supporting children in developing “low-tech prototypes” (Yip et al., 2013) of their designs by creating a visual story of a superhero robot using giant sticky-notes and markers. Researchers also supported children in realizing their fantastical designs (Guha et al., 2013) with the available robotics ML tools. Thus, through our iterative design process with children, we learned how to design mostly open systems in terms of allowing for daily reflective discussions, flexible ML technical tools, and flexible social conditions to reflect on social, ethical, and political ideas around ML but also around the design of the activities. However, these open systems were unsuccessful if the young learner had to fully control their own learning, likely overwhelmed with the amount of information and choice offered (Kirschner & van Merriënboer, 2013). Providing children autonomy in developmentally appropriate ways with adult guidance and support provided the conditions needed for meta-design.

Although we provided a particular set of tools, we discovered that children took advantage of the flexible technical conditions to create artifacts that went beyond the adult designer’s expectations. For example, Kendall created an ML robot that could teach underprivileged children, mitigated bias in her machine, and aligned her project to her interests and personal values. In another example, the Super Phenomenal Gamers team created a robot that promoted their YouTube channel, recorded their robot creation process, and posted these videos to promote their channel further. Before Bianca, Eric, and Ian found a way to integrate competitive gaming into their robot construction, they marginally participated in the program. Once they saw the program’s benefits and that they could augment their everyday tasks (Fischer, 2011), they became fully engaged in designing a ML robot. However, this success story had its limitations. The Super Phenomenal Gamers did not engage critically with ML, which was not aligned with the researchers’ goals. These findings speak to the multiplicity of voices that emerge during open systems in the meta-design of CML educational activities, which introduces complexities and tensions. In this case, if the researchers pushed their critical agenda with the Super Phenomenal Gamers, would the children have entirely disengaged from their ML project? These questions and the balancing of interests between children and researchers are essential to anticipate and engage with during meta-design around CML education.

In another form of meta-design interaction, researchers created flexible social conditions and engaged with youth in daily discussions about tools and the pedagogical activities. These discussions created an open culture of reflection that lowered the barriers to sharing design suggestions and allowed children to see that changes to the program and tools were indeed possible while the program was occurring (Fischer, 2011). However, when discussing the sociopolitical aspects of ML, the open culture of reflection became uncomfortable and unsafe for some of the youth who were part of the group being discriminated against by ML. Some youth felt passionate about discussing discrimination, but for some, the conditions did not facilitate this form of discussion. For example, Justin withdrew from the discussion when insensitive language was used and never addressed, and other children withdrew from the content around anti-Black racism as well. These examples suggest that although it is essential to engage youth in sociopolitical reflections at early ages (Ayón, 2016; Boutte & Muller, 2018), much preparation and training must go into developing such open systems for critical discussions around ML technologies. This preparation must go beyond consulting with community partners (Lee et al., 2021) to include how to respond compassionately and sensitively to children, how to allow them to opt out of discussions safely, and, more generally how to build trust among partners in politicized contexts in technology education (Vakil et al., 2016).

Although asking children to co-design the CML learning environment while simultaneously participating in the learning environment was fundamentally messy, conceptualizing the roles of children and adults as meta-designers and incorporating cooperative inquiry techniques benefited the design of the program and engaged more children in critical thinking around ML as the program continued and was re-designed. A cooperative inquiry approach provided appropriate tools for children to participate as design partners where their expertise and curiosity were valued. All in all, the children's projects surprised and inspired the researchers to redesign the activities during the program and for the future. In subsequent implementations, we will encourage children to incorporate digital media of their own choice and promote the idea of videoing and narrating their work. We will also encourage youth to think more critically about their designs. In addition to redesigning activities, we will also update the robot's functionality based on the children's desires, such as adding music or adding more crafting opportunities.

6. Conclusion

This study integrated cooperative inquiry techniques with meta-design approaches to describe how adults and children collaborated as meta-design partners in a CML program and simultaneously engaged in the program as learners. We argue that conceptualizing adults and children as meta-designers is a practical approach for collaborative design and ML applications for social good. The exploration presented in this study is just one example of the multiple possible approaches towards engaging youth early on in their education about machine learning concepts and how to think critically about the social, ethical, and political issues around modern large-scale algorithm deployment. Educational research explorations are crucial for breaking the harmful tradition of technology development and consumption without a critical lens.

Acknowledgment

Thank you to the lovely children at the local after-school centers for being our meta-design partners and developing this CML program. This work is funded in part by the National Science Foundation (ECR-2024965). The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

Ali, S., Payne, B. H., Williams, R., Park, H. W., & Breazeal, C. (2019). Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. *International Workshop on Education in Artificial Intelligence K-12 (EDUAI '19)* (pp. 1-4).

Arastoopour Irgens, G., Vega, H., Adisa, I., & Bailey, C. (2022). Characterizing children's conceptual knowledge and computational practices in a critical machine learning educational program. *International Journal of Child-Computer Interaction, 34*, 100541. https://doi.org/10.1016/j.ijcci.2022.100541

Ayón, C. (2016). Talking to Latino children about race, inequality, and discrimination: Raising families in an anti-immigrant political environment. *Journal of the Society for Social Work and Research, 7*(3), 449–477. https://doi.org/10.1086/686929

Bailey, C. S., Adisa, I. O., Vega, H., & Arastoopour Irgens, G. (2021). Cognitive, affective, and politicized trust in a community youth program: A Participatory design research project. In *2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT)* (pp. 1-2). IEEE. https://doi.org/10.1109/RESPECT51740.2021.9620658

Boucher, P. (2020). *Artificial intelligence: How does it work, why does it matter, and what can we do about it?* EPRS: European Parliamentary Research Service.

Boutte, G., & Muller, M. (2018). Engaging children in conversations about oppression using children's literature. *Talking Points, 30*(1), 2–9.

Braden, E. G. (2019). Do you have papers?: Latinx third graders analyze immigration policy through critical multicultural literature. *Children's Literature in Education, 50*(4), 464–480. https://doi.org/10.1007/s10583-018-9359-5

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, in *Proceedings of Machine Learning Research 81*, 77-91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Cumbo, B. J., Eriksson, E., & Iversen, O. S. (2019). The "Least-adult" role in participatory design with children. *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 73–84. https://doi.org/10.1145/3369457.3369464

Data Science for Social Good. (2021). *Data science for social good*. https://www.datascienceforsocialgood.org/

DiSalvo, B., & DesPortes, K. (2017). Participatory design for value-driven learning. In B. DiSalvo, J. Yip, E. Bonsignore, & C. DiSalvo (Eds.), *Participatory Design for Learning* (1st ed., pp. 175–188). Routledge. https://doi.org/10.4324/9781315630830-17

Druin, A. (1999). Cooperative inquiry: New technologies for children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 592–599). https://doi.org/10.1145/302979.303166

Ehn, P. (2008). Participation in design things. *Proceedings Participatory Design Conference (PDC)* (pp. 92–101). ACM Digital Library.

Eubanks, V. (2018). *Automating inequality. How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Fain, J. G. (2008). "Um, They weren't thinking about their thinking": Children's talk about issues of oppression. *Multicultural Perspectives, 10*(4), 201–208. https://doi.org/10.1080/15210960802526102

Fischer, G. (2011). Beyond interaction: Meta-design and cultures of participation. *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (pp. 112–121). https://doi.org/10.1145/2071536.2071553

Fischer, G. (2021). End-user development: Empowering stakeholders with artificial intelligence, meta-design, and cultures of participation. In D. Fogli, D. Tetteroo, B. R. Barricelli, S. Borsci, P. Markopoulos, & G. A. Papadopoulos (Eds.), *End-User Development* (Vol. 12724, pp. 3–16). Springer International Publishing. https://doi.org/10.1007/978-3-030-79840-6_1

Fischer, G., Giaccardi, E., Ye, Y., Sutcliffe, A. G., & Mehandjiev, N. (2004). Meta-design: A Manifesto for end-user development. *Communications of the ACM, 47*(9), 33–37. https://doi.org/10.1145/1015864.1015884

Fischer, G., & Scharff, E. (2000). Meta-design: design for designers. *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 396-405). https://doi.org/10.1145/347642.347798

Freire, P. (1970). *Pedagogy of the oppressed* (30th anniversary ed). Continuum.

Giroux, H. A. (1985). Critical pedagogy, cultural politics and the discourse of experience. *Journal of Education, 167*(2), 22–41. https://doi.org/10.1177/002205748516700204

Glesne, C. (2014). *Becoming qualitative researchers: An Introduction* (5th ed.). Pearson.

Guha, M. L., Druin, A., & Fails, J. A. (2013). Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction, 1*(1), 14–23. https://doi.org/10.1016/j.ijcci.2012.08.003

Jenkins, H. (2006). *Fans, bloggers, and gamers: Exploring participatory culture*. NYU Press.

Kantayya, S. (Director). (2020). *Coded bias* [Documentary]. 7th Empire Media. https://www.codedbias.com/

Kersten, J. M. (2006). Why's everyone white? Moving toward critical pedagogy in an elementary classroom. *Journal of Urban Learning, Teaching, and Research, 2*, 35–44.

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist, 48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Lee, I., Ali, S., Zhang, H., DiPaola, D., & Breazeal, C. (2021). Developing middle school students' AI literacy. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 191–197). https://doi.org/10.1145/3408877.3432513

Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism* (1st ed.). NYU Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown.

Pangrazio, L., & Selwyn, N. (2019). "Personal data literacies": A Critical literacies approach to enhancing understandings of personal digital data. *New Media & Society, 21*(2), 419-437. https://doi.org/10.1177/1461444818799523

Paris, D. (2012). Culturally sustaining pedagogy: A Needed change in stance, terminology, and practice. *Educational Researcher, 41*(3), 93–97. https://doi.org/10.3102/0013189X12441244

Paris, D. (2021). Culturally sustaining pedagogies and our futures. *The Educational Forum, 85*(4), 364–376. https://doi.org/10.1080/00131725.2021.1957634

Robertson, T., & Simonsen, J. (2013). Participatory design: An Introduction. In *Routledge International Handbook of Participatory Design* (pp. 1-17). Routledge.

Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education, 96*(4), 573–599. https://doi.org/10.1002/sce.21014

Samoili, S., Cobo, M. L., Delipetrev, B., Martinez-Plumed, F., Gomez, E., & Prato, G. D. (2021). AI watch. Defining Artificial Intelligence 2.0. Towards an operational definition and taxonomy of AI for the AI landscape. In *JRC Working Papers (No. JRC126426; JRC Working Papers). Joint Research Centre (Seville site)*. https://ideas.repec.org/p/ipt/iptwpa/jrc126426.html

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In *The SAGE Handbook of Qualitative Research in Psychology* (Vol. 2, pp. 17–37). Sage.

Vakil, S., McKinney de Royston, M., Suad Nasir, N., & Kirshner, B. (2016). Rethinking race and power in design-based research: reflections from the field. *Cognition and Instruction, 34*(3), 194–209. https://doi.org/10.1080/07370008.2016.1169817

Williams, R., Park, H. W., & Breazeal, C. (2019). A is for artificial intelligence: The Impact of artificial intelligence activities on young children's perceptions of robots. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–11). https://doi.org/10.1145/3290605.3300677

Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies, 129*, 41–54. https://doi.org/10.1016/j.ijhcs.2019.03.006

Yip, J., Clegg, T., Bonsignore, E., Gelderblom, H., Rhodes, E., & Druin, A. (2013). Brownies or bags-of-stuff?: Domain expertise in cooperative inquiry with children. *Proceedings of the 12th International Conference on Interaction Design and Children*, 201–210. https://doi.org/10.1145/2485760.2485763

Appendix A

Critical Data Literacies Project

Semi-structured Cognitive Interview Protocol

Introduction:

A. Obtain consent and answer questions about the study (if this is the first meeting and consent form has not been signed)
B. Review the study procedures with the participant
C. Verify permission to record the interview using a digital voice recorder/camera (are you okay with recording the interview?)
D. Say the participant's name and the date at the beginning of the recording
E. Ask the following questions:

Questions:

1. Can you show me your Teachable Machine?
    1. Walk me through your TM. How does it work?
2. How did you come up with ideas for your TM?
    1. Did you include any of your interests?
    2. How did you start your TM?
    3. What steps did you take?
    4. What was important for you in this stage? How about in the next stages?
3. Once you decided ________, how did you move on? What did you do?
    1. Why did you _______?
    2. How did you_______? / How did you learn to _______?
4. I noticed that you ________, tell me more about it.
5. Were there any challenges in _______?
    1. How did you work them out?
6. How do you feel about your process of creating your TM?
    1. What was something you enjoyed about it?
    2. What was something that you did not like about it?
7. How did you feel about the results of your machine?
    1. Did you share it with anyone else and if so, what did they think?
    2. Do you think anyone else would like to use your machine? Who?
    3. Is there anything you would change about it if you were to do this again?
8. If we were to come back in the summer or next year, what sort of things would you like to work on with us?

Louie, J., Stiles, J., Fagan, E., Chance, B., & Roy, S. (2022). Building toward critical data literacy with investigations of income inequality. *Educational Technology & Society, 25* (4), 142-163.

Building toward Critical Data Literacy with Investigations of Income Inequality

Josephine Louie[1*], Jennifer Stiles[1], Emily Fagan[1], Beth Chance[2] and Soma Roy[2]
[1]Education Development Center, USA // [2]California Polytechnic State University, USA // jlouie@edc.org // jstiles@edc.org // efagan@edc.org // bchance@calpoly.edu // soroy@calpoly.edu
[*]Corresponding author

ABSTRACT: To promote understanding of and interest in working with data among diverse student populations, we developed and studied a high school mathematics curriculum module that examines income inequality in the United States. Designed as a multi-week set of applied data investigations, the module supports student analyses of income inequality using U.S. Census Bureau microdata and the online data analysis tool the Common Online Data Analysis Platform (CODAP). Pre- and post-module data show that use of this module was associated with statistically significant growth in students' understanding of fundamental data concepts and individual interests in statistics and data analysis, with small to moderate effect sizes. Student survey responses and interview data from students and teachers suggest that the topic of income inequality, features within CODAP, the use of person-level data, and opportunities to engage in multivariable thinking helped to support critical data literacy and its foundations among participating students. We describe our definitions of *data literacy* and *critical data literacy* and discuss curriculum strategies to develop them.

Keywords: K–12 education, Intercultural competence, 21st century skills, Inquiry learning

1. Introduction

In a world steeped in data, many voices have called on schools to strengthen students' data literacy. Researchers argue that a wide range of data literacy skills are needed to meet rapidly growing workforce demands (Henke et al., 2016; Manyika et al., 2011) and to participate in modern civic life (Bargagliotti et al., 2020; Wolff et al., 2016). As ever more aspects of our lives become captured in data, additional voices have called for greater critical data literacy. These calls emanate from concerns about the potential harms to individuals and society when powerful groups collect vast amounts of our personal data and shape its narratives (e.g., D'Ignazio & Klein, 2020; Raffaghelli, 2020). To equip all individuals with the skills required to thrive in today's world, K–12 education needs scalable research-based strategies to strengthen and support diverse student populations in developing critical data literacy. This article describes a curriculum intervention developed and implemented in U.S. high school non-Advanced Placement (AP) mathematics classes to study one such strategy.

2. Conceptual approach

2.1. Defining data literacy and critical data literacy

For decades, statistics educators have worked to define and advance skills to make sense of data (Donoho, 2017; Rubin, 2020). Like Gould (2017) and Weiland (2017), we consider data literacy to rest on a foundation of statistical literacy. This type of literacy involves the abilities to interpret, assess, and communicate understandings of data from our everyday lives (Gal, 2002). It also includes basic fluency with the process of data investigation. As articulated by leading statistics educators, this process involves four iterative steps: (1) formulating questions that can be answered with data, (2) assembling data to address one's questions, (3) using statistical and other tools to analyze the data, and (4) interpreting results to address one's original questions (Bargagliotti et al., 2020). We also associate data literacy with a disposition to interrogate each step within the data investigation process (Bargagliotti et al., 2020). Data literate individuals carry a habit of mind (Cuoco et al., 1996; Finzer, 2013) that routinely asks about the provenance or origins of the data; who defined, measured, and collected the data; what tools or methods were used to analyze the data; and the degree to which data interpretations or conclusions are valid. Furthermore, we consider data literacy to include an ability to display multivariable thinking. This type of thinking recognizes that the relationship between two variables may not be as it initially seems given the possible effects of other interacting or confounding variables. Because real-world data and phenomena are typically multivariable, all learners should develop a capacity for multivariable thinking (Bargagliotti et al., 2020; Engel, 2016; Ridgway, 2015).

Building on this foundation, we consider critical data literacy to incorporate the skills of data literacy along with ethical and sociopolitical perspectives toward data. Digital technologies today allow actors to collect massive amounts of personal data to predict our behaviors and to target us with messages that can shape our social perceptions and actions (D'Ignazio & Klein, 2020; Noble, 2018; O'Neil, 2016; Pangrazio & Selwyn, 2019). Critical data literacy requires ethical consideration of whether and how to collect and use data from others and to reduce potential harm to individual privacy and autonomy (Baumer et al., 2022; National Academies of Sciences, Engineering, and Medicine, 2018). It also involves an awareness of how data can be used to perpetuate or aggravate unequal power relations in society, such as among groups that differ by race, sex, or economic class (e.g., Bhargava et al., 2015; D'Ignazio & Klein, 2020; Noble, 2018; Philip et al., 2016). People with critical data literacy not only scrutinize data for biases and possible misdirection, they also use data to understand and challenge forms of social oppression and to work toward more equitable social outcomes (Bhargava et al., 2015; D'Ignazio & Klein, 2020; Louie, 2022; Pangrazio & Selwyn, 2019).

2.2. Building toward critical data literacy

Prior research points to the types of classroom environments that are most likely to help students build foundational data literacy skills. In such environments, students explore motivating questions with real data, engage actively in the four steps of the data investigation process, use digital tools to help visualize and analyze data, and participate in rich classroom discussions to make sense of data (Bargagliotti et al., 2020; Berry et al., 2020; Chance et al., 2007; Garfield & Ben-Zvi, 2008). To help students recognize how data sets are constructed and their limitations, scholars suggest that students engage in primary data collection or examine secondary data collected by others in familiar contexts (e.g., Enyedy & Mukhopadhyay, 2007; Lee & Wilkerson, 2018; Rubel et al., 2016; Van Wart et al., 2020).

Scholarship on teaching mathematics for social justice, or TMSJ (Berry et al., 2020), offers ideas for fostering critical data literacy. TMSJ seeks to build students' capacities to recognize and counter social injustices with mathematics and data. Gutstein (2003, 2006), drawing on the ideas of Freire (1970), describes teaching for social justice as having three primary goals: (1) developing students' social and political consciousness, (2) building agency to affect social change, and (3) fostering positive social and cultural identities. Bringing this approach to mathematics education, he describes TMSJ as helping students to "read" and "write" the world with mathematics—that is, to discern and devise solutions to pressing social problems with quantitative data and reasoning (Gutstein, 2003; Gutstein, 2006). TMSJ invites students to investigate complex, real-life issues using data that reveals the nature and scope of unequal socioeconomic outcomes among groups that differ by race, class, sex, or other social characteristics and to raise questions about the factors that contribute to these outcomes. Scholars of TMSJ have organized instruction around projects examining issues such as regressive utility rates that cause low-income families to pay higher electricity rates than wealthy energy users (Frankenstein, 2009; Frankenstein, 2013); racial disparities in police stops and access to housing markets (Gutstein, 2003; Gutstein, 2013); and disparate racial, class, and spatial patterns in public lottery participation (Rubel et al., 2016).

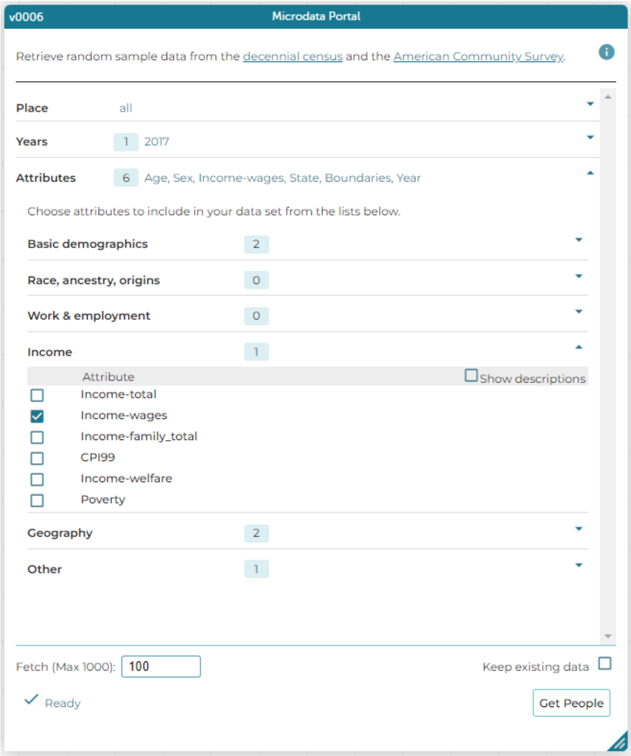
Another key feature of TMSJ is the use of culturally relevant pedagogy to advance equitable learning opportunities among diverse student populations (Berry et al., 2020). Culturally relevant pedagogy values and incorporates students' cultural backgrounds and lived experiences in classroom instruction, both to facilitate students' academic learning and to strengthen their identities (Gay, 2002; Gay, 2013; Ladson-Billings, 1995). It also emphasizes high expectations and standards-based instruction for students from historically marginalized groups to help them achieve academically and to access social and economic power (Berry et al., 2020; Delpit, 1988; Ladson-Billings, 1995). Research has found that culturally relevant pedagogy can lead to improved academic engagement and learning outcomes among diverse student populations, particularly those whom educators label as "at risk" (Aronson & Laughter, 2016; Dee & Penner, 2017). Because the goals of critical data literacy, TMSJ, and culturally relevant pedagogy are so highly aligned, employing strategies of TMSJ and culturally relevant pedagogy may be an effective way to promote critical data literacy among diverse populations of students.

3. Data literacy modules focused on social justice issues

Drawing on these ideas, we developed a set of high school curriculum modules that consist of data investigations examining social justice issues. A primary goal of our work has been to promote understanding of and interests in fundamental statistical thinking practices (i.e., data literacy) among diverse student populations with high proportions of learners from historically marginalized groups. A second goal has been to explore the extent to which module use may advance aspects of students' critical data literacy, focusing on two outcomes that Gutstein identified as central to social justice learning: acquiring social and political consciousness and gaining a sense of social agency. Designed as a series of lessons that teachers facilitate over 15 one-hour class periods, the modules guide students in exploring social and economic questions of direct relevance to themselves, their families, and their communities. Students address these questions by analyzing large-scale person-level microdata from the American Community Survey (ACS) and the U.S. decennial census, using the free and browser-based Common Online Data Analysis Platform (CODAP).

In one module, *Investigating Income Inequality in the U.S.* (the "Income Inequality module"; full module materials can be found at https://go.edc.org/ussdata), students work through seven lessons and a final team data investigation, discussing different forms of income inequality, its scope, and its causes. They begin in Lesson 1 by grappling with the following question: What is income inequality, and when might it be a concern for society? With teacher facilitation, they discuss how they think total income in the United States is distributed among lower- and higher-income groups, how this distribution has changed over time, the questions these topics raise for them, and whether they can answer their questions with data. In Lesson 2, students consider the provenance, strengths, and limitations of data from the U.S. decennial census and the ACS. They learn about the U.S. federal government's stated purposes and uses of these data sources, as well as the types of data that are collected (e.g., multiple measures of income but not wealth) by reviewing and attempting to answer ACS questionnaire items themselves. Students discuss the reasons why people may skip or have difficulty answering specific questions and the implications of missing or inaccurate responses.

Figure 1. The U.S. decennial census and ACS Microdata Portal in CODAP



Note. The Microdata Portal within CODAP provides access to person-level data from the U.S. decennial census for 11 decades (from 1860 to 2010) and from the ACS for 2017. Users can draw random samples of individuals from the United States as a whole as well as from each of the 50 states. They can draw samples of up to 1,000 people at a time, with the capacity to add cases to create larger samples. The data source is IPUMS-USA, University of Minnesota, www.ipums.org.

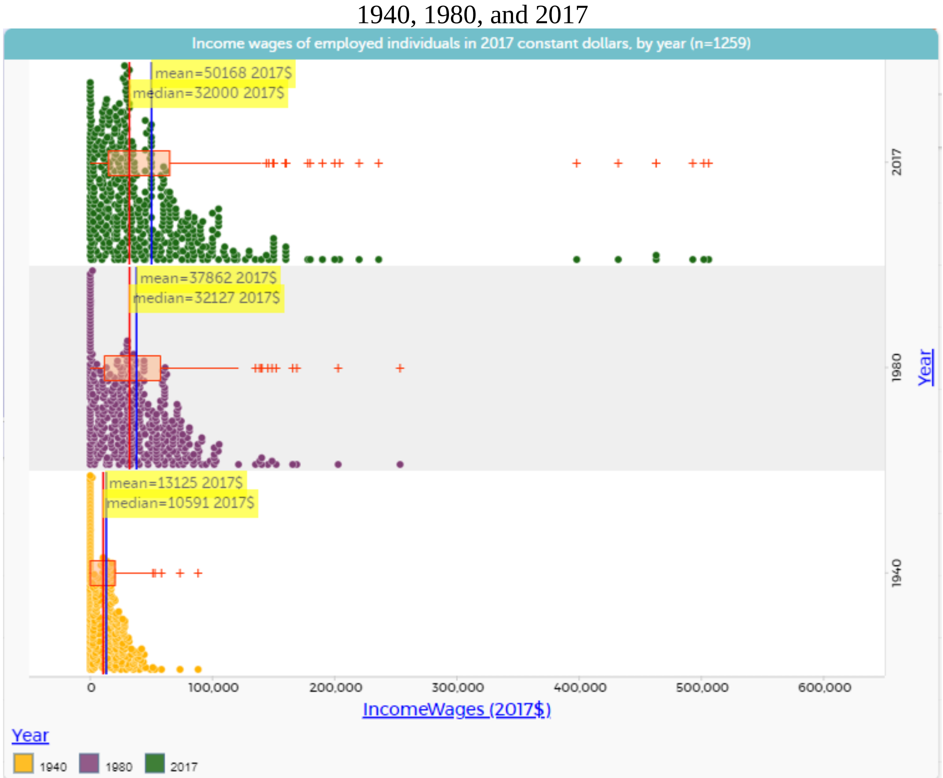
Subsequent lessons ask students to investigate questions about individual and aggregate incomes of U.S. wage earners, using data drawn from a U.S. census and ACS Microdata Portal that we developed as a permanent plug-in tool for CODAP (Figure 1). We provided students with access to person-level microdata with over three dozen individual attributes (or variables) for exploration, under the hypothesis that students would be better able to relate to the people in the data if they could see person-level records in their data sets. A person-level data set would also allow students the flexibility to examine a wide range of relationships among multiple variables of their own choosing, such as the relationship between income and education when controlling for race. Prior studies involving map-based investigations of social issues with large-scale social data (e.g., Kahn, 2020; Radinsky et al., 2014; Rubel et al., 2016) have allowed students to examine social and economic outcomes based on neighborhoods or higher units of geography, but not based on individuals as the unit of analysis.

In Lessons 3–5, students investigate how income equality in the United States has changed over time. They begin by creating data tables that contain random samples of individuals from the U.S. population in 2017, with information about each person's reported annual wages and other characteristics (such as age, sex, race/ethnicity, and occupation). They then create dot plots of the wage distribution by dragging and dropping the wage attribute from their data table to a graphing window and watching data cases populate the graph dynamically. Students can hover over or click on individual dots to see an individual's specific attributes in a pop-up window or in the linked data table (Figure 2).

Figure 2. A CODAP data table and dot plot of annual income from wages for a random sample of employed U.S. individuals in 2017

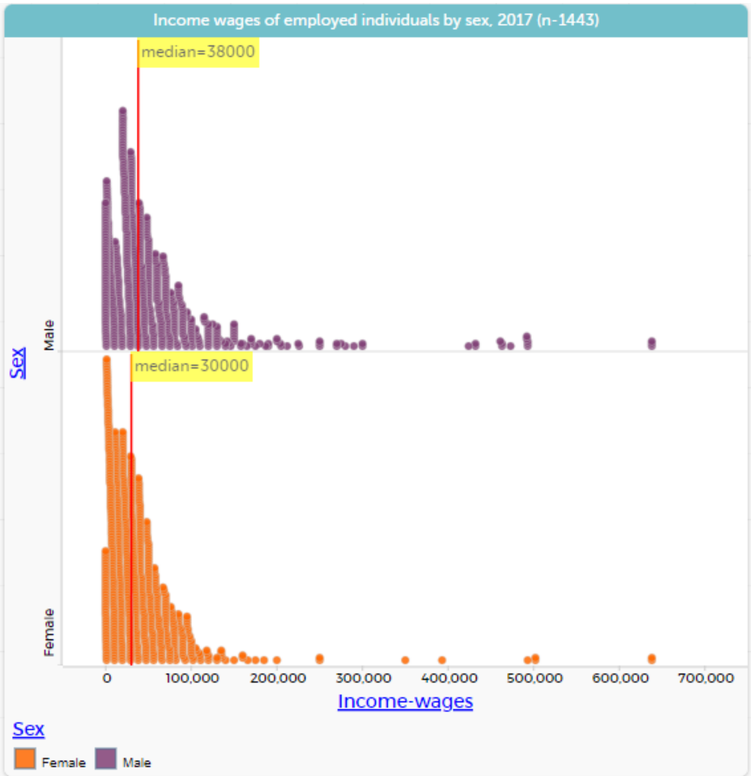
Note. CODAP allows users to hover over or click on individual cases in the graph to highlight case information within the graph itself and in the linked data table. A toolbar provides functions that allow users to display measures such as the mean and median of the distribution.

Figure 3. CODAP dot plots of annual income from wages for random samples of employed U.S. individuals in 1940, 1980, and 2017

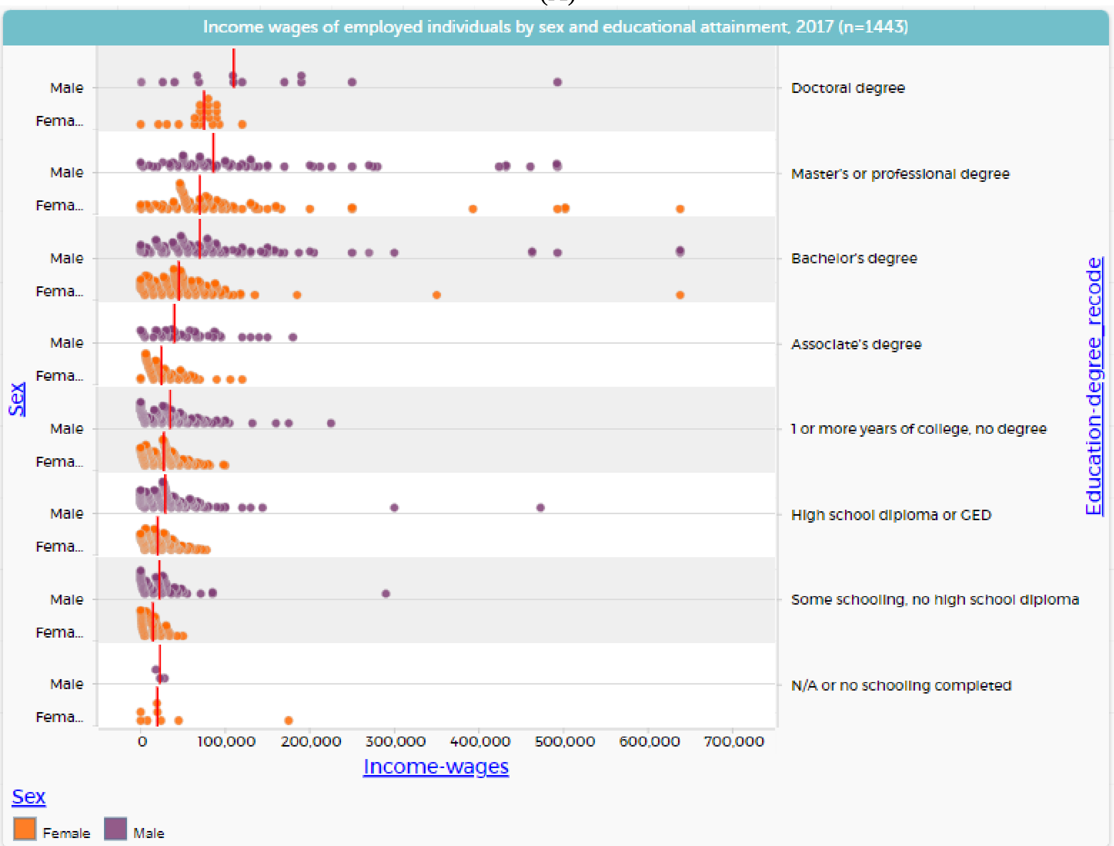


Note. Students analyze annual income from wages in each decade as measured in constant 2017 dollars.

Figure 4. Graphs of wages by sex (A), then by educational attainment (B), for a random sample of employed U.S. individuals in 2017



(A)



(B)

The dot plots in CODAP allow students to see the shape of the entire wage distribution, with the bulk of individuals reporting wages at the low end of the income spectrum and a small number of individuals reporting wages at the upper end. The strong skew of the wage distribution and the CODAP tools help students build a conceptual understanding of the difference between the mean and the median—two concepts that students traditionally learn to compute without a deeper grasp of their meaning or when to use them (Konold & Higgins, 2003). In CODAP, students can turn on visual indicators of where the mean and median fall within the distribution; they can also hold and drag an outlying high-wage case further to the right to see the mean of the wage distribution increase while the median does not change. Students discuss whether the mean or median value is a better representation of the sample's "typical" wage, given the skew in the data. They also discuss how the distance between the two measures can indicate the presence of a small number of extreme cases "pulling" the mean away from the median, thus providing one way to measure the degree of income inequality in the sample.

Students use these ideas to examine how wages have diverged between lower- and higher-income earners since 1940 (Figure 3). They examine the changing shape, spread, whisker length, and number of statistical outliers associated with box plots superimposed on the data. They quantify growing gaps among higher- and lower-wage earners by calculating the ratio of annual wages of individuals at the top and bottom quartiles each decade. Furthermore, in a class activity, they build an intuition for the margin of error for their sample estimates by using the Microdata Portal to select multiple samples of varying sizes (starting from n = 100 and growing by increments until they reach 1,000 cases) and observing the level of variation across their estimates.

In Lesson 6, students turn to examine income inequality between male and female employees in the United States. Using data drawn from the Microdata Portal, students make graphs (e.g., Figure 4, graph A) to compare measures of center and variability for the wage distributions of male and female employees. In Lesson 7, after speculating about possible causes behind differences in wages by sex, they examine whether the gap in median wages for the two groups can be explained by education. They explore this question with a graph "splitting" function, which allows students to visualize how the relationship between two variables may change after adjusting for a third variable (Figure 4, graph B). Once they discover that the wage gap persists after controlling for education, students discuss if other individual attributes may explain the gap or if sex discrimination may be at play. In a final team data investigation, students collaborate in groups to choose and investigate another variable that may explain the gap in wages by sex. They enact, document, and present findings from the entire data investigation process. Through these activities, they deepen their abilities to describe and compare quantitative distributions using measures of center and variability—key practices in the Common Core State Standards for Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) for secondary grade levels. They also build abilities to reason with multivariable data, a skill that should be but is not currently emphasized in K–12 education (e.g., Engel, 2016).

4. Module iterations and study research questions

With input from both high school mathematics and social studies teachers, we developed an "alpha" iteration of the Income Inequality module, which five high school mathematics teachers implemented with their students during the 2018-2019 academic year. We then revised the materials and tested a "beta" iteration of the module with seven high school mathematics teachers and their students during the 2019-2020 academic year. Testing of this module ended just at the onset of the COVID-19 pandemic.

The questions that drive this study derive from the project goals described earlier. We also aim to expand on existing research. Prior studies of mathematics or data-focused interventions with a social justice or critical literacy orientation have involved smaller-scale classroom or summer session case studies (e.g., Enyedy & Mukopadhya, 2007; Van Wart et al., 2020; Rubel et al., 2016) and research in which study leaders enacted the interventions themselves (e.g., Gutstein, 2003; Brantlinger, 2013). In contrast, we sought from the start to recruit multiple teachers to test our modules in authentic classroom settings with high proportions of students from historically marginalized groups, where we (as researchers, curriculum developers, and higher education faculty members) were not involved in direct module implementation. We also wanted to examine whether module use is associated with quantitative gains in students' learning and interest outcomes to provide evidence of promise for a curriculum approach that teachers could implement and test at a larger scale. We sought to address the following research questions:

- RQ1. To what extent do students in participating classrooms show increased understanding of and interests in data literacy after completing the Income Inequality module?
- RQ2. What aspects of the module appear to promote participating students' understanding of and interests in data literacy?
- RQ3. In what ways do participating students demonstrate critical data literacy or social justice learning outcomes after completing the Income Inequality module?

5. Methods

5.1. Study team, sample, and teacher preparation

The leaders of the project are an educational researcher (Asian American, female) with a background in social policy research and two university-level statistics educators (one White American, female; one South Asian, female). The other team members are a senior mathematics curriculum writer and a research associate (each White American, female). During the early years of the project, a senior scientist/technology developer (White American, male) was heavily involved in development of CODAP tools for the modules. All members of the project team participated in different aspects of module development and classroom research.

We recruited seven high school mathematics teachers from six cities in a northeast U.S. state to implement the Income Inequality module in non-Advanced Placement (AP) mathematics classes in fall 2019 and early 2020. We targeted public high schools with high proportions of students from Black, Latinx, and low-income families, sharing a letter seeking teachers "to help test and inform the development of a set of applied data investigations in high school non-AP mathematics classes" to support students in "investigating U.S. socioeconomic trends and questions related to social justice, using large-scale U.S. population and economic data sets from the U.S. Census Bureau and other government agencies." Unlike other studies (e.g., Kokka, 2020), teachers did not have to profess a social justice orientation in their instruction to participate. In screening interviews, teachers whom we accepted in our study expressed a strong interest in our curriculum approach, particularly its project-based orientation, use of real data, and investigation of real-life issues.

Table 1. Characteristics of study participants

A. Teacher, classroom, and student characteristics

Teachers[1]			Classrooms[2]		Students[3]				
Name	Highest degree	Years teaching math/statistics	Course sections (#)	Students per section (#)	Age range	Black (%)	Latinx (%)	White (%)	Work to support self or family (%)
Eve	MA/CAGS	9/2	5	20	17–19	17	57	18	70
Rachel	MA/CAGS	22/10	1	30	16–19	10	50	25	75
Sasha	BA	3/2	2	10–15	17–19	30	30	20	60–70
Julie	BS	6/6	2	24	17–19	20	35	40	70
Anne	MA	16/14	1	25	16–18	20	30	40	80
Will	BA	9/5	2	25–30	16–19	20	40	25	50
Bella	MA	15/3	1	10–15	16–23	70	20	10	70

B. School characteristics[4]

School name	Teacher participant(s)	Enrollment (#)	Black (%)	Latinx (%)	White (%)	First Language not English (%)	Low income (%)
Elm	Eve	2,000	20	53	21	60	49
Rose	Rachel	2,020	4	58	31	67	45
Spruce	Sasha	1,230	13	47	32	58	45
Wood	Julie, Anne	1,650	10	45	39	51	38
Beech	Will	1,110	19	42	31	48	61
Ash	Bella	410	48	43	6	35	65

Note. [1]Data are from teacher self-report for the 2019-2020 school year. Teacher names are pseudonyms. All teachers self-identify as White and female except for Will (male) and Sasha (non-binary). All majored in mathematics as undergraduates except for Julie (engineering) and Bella (history). [2]Data are from teachers' reporting before or at the start of the 2019-2020 school year. [3]Data are from teachers' retrospective reporting during the 2021-2022 school year. [4]Data are from the state department of education for the 2019-2020 academic year. School names are pseudonyms. Enrollment figures are rounded to the nearest 10. All schools are public, include grades 9–12, and are in separate cities in the Northeast.

Table 1 provides background information on the seven participating teachers, as well as on their classrooms, students, and school contexts. All teachers taught grade 12 non-AP courses in statistics or data analysis and had the flexibility to incorporate multi-week curriculum modules into their course schedules. None of the teachers had prior experience working with large-scale data or with data analysis tools other than spreadsheet programs such as Microsoft Excel. Between 55% and 91% of students in the participating schools were Black or Latinx, and teachers estimated that 50%–80% of their students held jobs to help support themselves and their families. In each school, the percentages of students classified as low-income exceeded statewide percentages. The Ash school has a particularly high-needs population: It is an alternative public school that serves students who have had long-standing attendance issues or have not been successful in other schools.

We provided all participating teachers with an in-person seven-hour workshop in summer 2019 and a follow-up two-hour after-school session in the early fall to prepare them to implement the module with students. During these sessions, we provided a module overview, supported teachers in learning CODAP, and facilitated select lessons to allow teachers to experience the materials as learners. Three of the teachers (Rachel, Sasha, and Julie) had implemented the module's alpha iteration and provided data to inform revisions for the beta iteration. The remaining four teachers (Eve, Anne, Will, and Bella) were new to the project. During module implementation, project team members were available by phone and email to help teachers troubleshoot technology and other lesson questions.

5.2. Data sources and measures

To address RQ1, we invited all participating students to complete an online data literacy assessment and interest/affect survey immediately before starting and after completing the module. The assessment contained 19 multiple-choice items drawn primarily from item pools developed and validated by Jacobbe et al. (2014) and Garfield et al. (2006). We selected items to measure understanding in five domains that were aligned with the module's data literacy learning objectives: sampling and data collection, data representation, measures of center, measures of variability, and multivariable thinking (see Appendix 1). The interest/affect survey contained 23 seven-point Likert-type items adapted from validated scales by Linnenbrink-Garcia et al. (2010) and Sproesser et al. (2016). We assembled scales measuring students' temporary situational interest, longer-term individual interest, self-concept, and perceptions of the value in statistics and data analysis (see Appendix 2, Table 7 for sample items and Louie et al. (2021) for more on scale constructs and development). Cronbach's alpha for each scale ranged from 0.73 to 0.92, suggesting acceptable to very good reliability (DeVellis, 2012).

To address RQ2 and RQ3, we drew upon several data sources. First, we collected student responses to an open-ended prompt in the post survey, which asked: *Tell us how much you liked the Income Inequality data lessons compared to your other statistics lessons. Please explain your reasons*. Second, we conducted five focus group interviews with n = 4–6 students per group at four schools (Elm, Rose, Wood, and Beech). The project's lead researcher conducted the semi-structured interviews asking questions such as the following: *In what ways were the Income Inequality data lessons different from your other work in this class? What aspects of the lessons were [most/least] interesting, and why? What do you think you learned from these lessons?* Each interview lasted 40–60 minutes and was audio recorded and transcribed for analysis. Third, we conducted 1- to 2-hour interviews with five teachers soon after they completed the module and again in early 2022 as an auxiliary source of data about students' responses to the module. All teacher interviews were conducted by videoconference and recorded and transcribed for analysis. Fourth, three project team members conducted classroom observations during module implementation, with five teachers observed twice and two teachers observed once. Observers recorded written field notes, documenting classroom environments and how teachers implemented and how students responded to module activities. Fifth, we collected online implementation logs from teachers for contextual information on how they facilitated each lesson.

5.3. Data analyses

We calculated student scores on the data literacy assessment as the total number of correct responses and on the interest/affect scales as the average rating among scale items. We compared students' pre- and post-module scores using paired *t*-tests. To analyze student responses to the open-ended prompt in the post-survey, two project researchers reviewed all student responses; developed a set of overarching codes under the general categories of "positive," "neutral," and "negative" responses to each module; and created a more detailed set of subcodes. For example, subcodes under the positive category included the focus on a relevant or real-life topic, the use of CODAP, and the ease of navigating the materials. Each researcher independently assigned codes to student responses and arrived through discussion at a consensus set of coded data.

To analyze the student focus group data, two project researchers developed a set of overarching codes identifying when students spoke about something they had learned or had found interesting in the module; subcodes identifying whether their learning or interest was related to data and statistics or the social justice topic; and subcodes about aspects of the module (e.g., CODAP, the pedagogical approach) that supported their learning or interest. A parallel set of codes described challenges to students' learning or interests. The two researchers independently coded common transcript excerpts, met to resolve disagreements and to refine code definitions, and iterated on this process until reaching 91% interrater agreement, after which one researcher coded the remaining data. The other researcher applied the same set of codes to the teacher interview data, focusing on teachers' descriptions of student learning and interest during the module. The project's lead researcher analyzed all coded statements to identify emergent themes by code, which were discussed with the project team and teachers for corroboration or refinement. Data from teacher implementation logs and classroom observations were also reviewed by the team to add context and to help support or qualify findings from student and teacher interview data.

6. Results

6.1. RQ1: Changes in student data literacy and interests

Based on total scores on the data literacy assessment, students' understandings of assessed statistical concepts grew between the start and end of the module, and the change was statistically significant at $p < .0001$ with a moderate effect size ($d = 0.43$; Table 2). Students showed the greatest growth in their understanding of measures of center ($p < .0001$, $d = 0.38$), multivariable thinking ($p < .0001$, $d = 0.35$), and data representation ($p < .0001$, $d = 0.19$). Growth was largest in Eve's classes ($d = 0.89$; Table 3), which also had the largest number of participants. Although we offered each student a $20 Amazon gift card to complete both pre- and post-module assessments, attrition levels ranged from high to extremely high and varied widely by teacher (from 13% to 79%).

Table 2. Students' pre- and post-module data literacy scores, total and by domain

	Pre mean (*SD*)	Post mean (*SD*)	Post - Pre mean (*SD*)	*p*-value	Effect size (*d*)
Total (19 items)	9.66 (2.67)	10.94 (3.16)	1.28 (2.99)	< .0001	0.43
Sampling & Data Collection (3 items)	1.53 (0.80)	1.57 (0.79)	0.04 (0.92)	.560	0.04
Data Representation (4 items)	2.91 (0.91)	3.13 (0.97)	0.22 (1.15)	.011	0.19
Measures of Center (5 items)	2.15 (1.08)	2.68 (1.28)	0.53 (1.41)	< .0001	0.38
Measures of Variability (4 items)	1.57 (0.91)	1.71 (0.97)	0.13 (1.19)	.144	0.11
Multivariable thinking (3 items)	1.50 (0.86)	1.86 (0.89)	0.36 (1.03)	< .0001	0.35

Note. Items were multiple choice and scored as 1 = correct, 0 = incorrect or missing, with $n = 180$ students who provided both pre- and post-module responses.

Table 3. Students' pre- and post-module data literacy scores, total scores by teacher, with attrition rates

	Pre mean (*SD*)	Post mean (*SD*)	Post - Pre mean (*SD*)	*p*-value	Effect size (*d*)	Pre *n*	Post *n*	Attrition (%)
All teachers	9.66 (2.67)	10.94 (3.16)	1.28 (2.99)	< 0.0001	0.43	277	180	35
Eve	9.27 (2.68)	12.05 (3.15)	2.79 (3.13)	< 0.0001	0.89	100	75	25
Rachel	10.14 (1.21)	11.14 (2.41)	1.00 (2.83)	0.386	0.35	32	7	78
Sasha	6.60 (2.41)	6.40 (2.30)	- 0.20 (1.92)	0.828	- 0.10	24	5	79
Julie	9.51 (2.50)	10.43 (2.96)	0.91 (2.01)	0.011	0.46	47	35	26
Anne	9.58 (1.83)	8.92 (2.97)	- 0.67 (2.96)	0.452	- 0.23	20	12	40
Will	10.66 (2.73)	10.37 (2.75)	- 0.29 (2.39)	0.438	- 0.12	47	41	13
Bella	11.00 (3.87)	11.80 (2.59)	0.80 (1.79)	0.374	0.45	7	5	29

Note. Post *n* represents the number of students who provided both pre- and post-module responses.

Table 4. Students' pre- and post-module interest and affect scores related to data literacy, by scale

	n^1	Pre mean (*SD*)		Post mean (*SD*)		Post - Pre mean (SD)		*p*-value	Effect size (*d*)
Situational interest (8 items)	192	4.95	(0.97)	4.98	(1.05)	0.03	(1.01)	.689	0.03
Individual interest (4 items)	194	3.81	(1.20)	4.11	(1.23)	0.31	(1.16)	.000	0.26
Self-concept (5 items)	195	4.71	(1.04)	4.72	(1.05)	0.01	(0.81)	.873	0.01
Perceptions of value (6 items)	189	4.96	(0.94)	5.08	(0.93)	0.12	(0.83)	.052	0.14

Note. A student's scale score is the average rating for each scale item, which could range from 0 to 7. [1]Sample size *n* represents the number of students who provided pre- and post-module responses for each scale.

Results from the interest/affect survey suggest that changes in students' temporary situational interest, self-concept, and perceptions of the value in statistics and data analysis between the start and end of the module were not statistically significant at $p < .05$, whereas growth in students' deeper individual interest in these domains was significant with a small effect size (p = .001, d = 0.25; see Table 4). The overall results primarily reflect strong growth in ratings among Eve's students (p = .000, d = 0.59). Like before, attrition rates and effect sizes for the interest/affect scales varied widely by teacher (Table 5).

Table 5. Students' pre- and post-module scores on the individual interest scale, with attrition rates

	Pre mean (*SD*)		Post mean (*SD*)		Post - Pre mean (*SD*)		*p*-value	Effect size (*d*)	Pre *n*	Post *n*	Attrition (%)
All teachers	3.81	(1.20)	4.11	(1.23)	0.31	(1.16)	.000	0.26	277	194	30
Eve	3.86	(1.20)	4.55	(1.20)	0.69	(1.17)	.000	0.59	100	76	24
Rachel	3.75	(1.04)	3.66	(1.17)	- 0.09	(1.53)	.868	- 0.06	32	8	75
Sasha	3.84	(0.96)	4.22	(0.71)	0.38	(0.73)	.191	0.51	24	8	67
Julie	3.30	(1.07)	3.49	(1.11)	0.18	(0.79)	.137	0.23	47	42	11
Anne	4.46	(1.29)	4.21	(1.07)	- 0.25	(1.29)	.481	- 0.19	20	14	30
Will	3.96	(1.20)	3.98	(1.22)	0.02	(1.23)	.900	0.02	47	41	13
Bella	4.30	(1.63)	4.15	(1.81)	- 0.15	(1.42)	.825	- 0.11	7	5	29

Note. Post *n* represents the number of students who provided both pre- and post-module responses.

Although student outcomes varied by teacher, classroom observations and teacher logs did not uncover systematic differences in how teachers implemented the module. An exception was with Eve, who tended to keep her students working together at the same pace while other teachers let small groups proceed at different paces. Teachers reported that they made few modifications to the materials and skipped few activities. There were two exceptions: Sasha skipped the lesson that examines wage distributions over time, and neither Sasha nor Eve implemented the final team data investigation due to time constraints. All teachers described problems with student attendance, with absenteeism most acute in Sasha's and Bella's classes. Attendance was so irregular in Bella's class that she required six weeks to complete the module; other teachers completed the module in three weeks or less.

6.2. RQ2: Aspects of the module that supported students' understanding of and interests in data literacy

Students' responses to an open-ended prompt in the post-module survey provide insight into aspects of the module that played a salient role in supporting students' understandings of and interests in statistics and data analysis. When asked to state how much they liked the module compared to their other statistics lessons, 86% of respondents shared positive comments about the module (e.g., "I liked the income inequality lessons a lot"), while 5% of the responses were neutral (e.g., "I don't really have an opinion"), and 13% offered comments that were negative ("not that much"; Table 6). In some cases, students responded with both positive and negative comments (e.g., "it was definitely better but still kinda boring to me").

Although many students did not provide reasons for their responses, a plurality of students indicated that they liked the module due to its focus on income inequality. About 32% of students who spoke positively of the module wrote that the topic of income inequality was "real," "relevant," or "relatable," and another 32% said the topic was "interesting." Some students said they liked examining an authentic problem that exists in their own lives, unlike the artificial problems or foreign contexts they encounter in textbooks. In a focus group interview, some of Will's students gave examples of these textbook problems, such as "comparing GPA to height," looking at "Olympic 100-meter sprint times over the past 25 years," and examining "the longevity of a cowboy." Students described these examples as "so boring," and one student asked: "What does this have to do with what we're gonna need?" In line with this last comment, other students suggested that they liked the module because the topic had practical value for their lives. A student from Eve's class wrote: "I actually like [the module] because we are learning about business money, wages… [and] I will like to take a business career some day." Similarly, one of Rachel's students wrote that the module "was the most interesting because it had real data that was useful for the future to see how I could make money." Other students suggested that they found the social justice aspect of the topic compelling. One of Eve's students wrote: "I really liked the income inequality data lessons a lot more than regular statistics lessons because income inequality is something that I find very interesting and I feel passionate about."

Another set of reasons that students gave for liking the Income Inequality module involved using CODAP. Working with this tool was a novel experience for students in each participating classroom. Only two teachers (Sasha and Bella) reported any type of student computer use in their statistics classes, where students occasionally used an online spreadsheet program (Google sheets) to analyze small data sets compiled from class surveys or online sources. The remaining teachers reported that their students used graphing calculators to enter data (typically no more than 40 cases) and to compute statistics by hand. Julie's typical class routine, which other teachers said they shared, was to discuss statistics concepts from a textbook and then to have students engage in "bookwork," that is, solve practice problems on paper with the aid of their calculators.

Table 6. Frequency and percent of coded student responses to the Income Inequality module[1]

Response codes	*n*	% of total[3]	% of subtotal[4]
Total student responses[2]	209	100	--
Positive response to the module	180	86	100
Module was "fun," "engaging," "interesting," or students "liked" or "enjoyed" it.	145		81
The topic of income inequality was "real," "relevant," or "relatable."	58		32
The topic of income or income inequality was interesting.	58		32
Enjoyed using CODAP, the interactive graphs, or the hands-on learning.	30		17
Learned or gained understanding, especially about society.	25		14
Found the lessons clear, well-structured, or easy to follow.	22		12
Not sure or neutral reactions to the module	10	5	--
Negative response to the module	28	13	100
Lessons were difficult or confusing.	6		21
Did not like working with CODAP.	3		11
Lessons or questions were too repetitive.	3		11

Note. [1]In the post-module survey, students responded to the following prompt: "Tell us how much you liked the Income Inequality data lessons compared to your other statistics lessons. Please explain your reasons." [2]Includes all students who provided a response to the open-ended prompt in the post-module survey. [3]Responses do not sum to 100% because some responses included both positive and negative comments. [4]Responses do not sum to 100% of each subtotal because coding categories are not mutually exclusive, and this table displays only those response codes that were applied to at least 10% of all positive or negative comments.

In this context, one of Eve's students wrote that the module "felt like an advanced lesson because we had to use our laptops," and another noted that "it was nice to work on a computer rather than stare at a white board." A classmate commented that using CODAP helped the module feel like "more of an interactive lesson where you are more engaged into the data and [you're] really doing and analyzing the data." Yet another of Eve's students explained that the ability to manipulate and visualize data in CODAP was engaging and supported her learning. She wrote:

> I feel that the income inequality data lessons provided more hands-on learning with the teacher and classmates. Making it very easy for everyone to learn and understand. As well as giving a very good visual representation of the data. I really like learning this way, which is why I've enjoyed this statistics class more than my other statistics classes I've taken before. *(Written survey response from one of Eve's students)*

Teachers indicated that they too thought that CODAP features helped to deepen students' understandings of key data concepts. Will stated in an interview that "CODAP and the hands-on activity…really started showing them what a skewed distribution does. And how mean and median change, why it's sometimes more appropriate to use [the] median." Bella shared that previously, she would simply tell her students that "you really only need 1,000 people to get good data about the whole U.S." With the Microdata Portal, however, students could "experientially see it in terms of the data that each group is generating and how little variation there is of a sample size of 1,000." For her, CODAP tools and activities "really helped build students' conceptual understanding" of central statistical ideas.

In addition, the use of person-level microdata, as well as the ability to see people represented as individual dots in a graph and as discrete records in a linked data table, helped students understand the structure of the data and connect to its human referents. A student in Will's class wrote in the post-module survey that "using the [CODAP] software made it easier to understand where data comes from." Teachers corroborated the value of person-level data as well as accompanying module activities. Both Julie and Rachel had implemented the alpha version of the module, when students had jumped immediately into making graphs with a random sample of 100 people from the ACS. In the beta version, students began by examining the ACS questionnaire and conducting a scavenger hunt to locate individuals with specific demographic characteristics in a CODAP table and a dot plot of 12 cases from the ACS. During a teacher focus group interview, Julie observed:

> [The] most powerful thing about this lesson was connecting it back to each dot on the graph representing a person and [students] being able to identify that—'Wait, that's me… That's my person.' And last year, we didn't have this activity and I realized later on in the modules that they could not connect the fact that each

> dot was a person. And so doing this up front made it a lot easier throughout the entire module. *(Comment from Julie during a teacher focus group interview)*

Rachel concurred, saying, "My students seemed to be pretty engaged…I just kept telling them this is a real person… [like] there is someone who lives in New Jersey... with this salary who is this age… They did like looking to find their person or people who were similar or dissimilar to them."

Another CODAP feature that students liked was the ability to explore relationships among multiple variables at the same time. Students expressed curiosity and motivation to uncover whether another variable such as education could explain the gap in median wages among male and female employees. CODAP allowed them to drag and drop a third variable into a graph of wages by sex and watch the graph split to illustrate new wage distributions by sex and the third variable. This CODAP feature helped students deepen their understanding of how graphs can be organized and structured. Some of the third variables that students chose to explore were a person's occupation, part- or full-time status, race/ethnicity, marital/family status, and age. A student from one of Julie's classes wrote in the post-module survey:

> I liked [the module] because we were able to investigate different attributes to try and figure out what causes the income gap between males and females... Also making different graphs to those different attributes was pretty cool to try and figure out so we have a better understanding on how to work and create graphs. *(Written survey response from one of Julie's students)*

During a focus group interview, a student from one of Julie's classes echoed these comments, noting that in their mathematics classes they normally examine only bivariate relationships and do not use tools that allow them to see data relationships change dynamically in multivariable contexts. The student explained:

> When we had the income and an attribute, and then we had the gender, we're not used to having so many different variables on the actual graph. Usually, it's just two on the graph and two on a separate graph. This way we could have one graph with all the things... It was cool watching how the data would in front of us change while we added an attribute. (*Focus group comment from one of Julie's students*)

6.3. RQ3: Students' critical data literacy or social justice learning outcomes

Although we did not embed explicit critical data literacy learning objectives in the module lessons, we were interested in exploring whether and in what ways students demonstrated outcomes that Gutstein (2003, 2006) describes as goals of social justice pedagogy. Based on both student and teacher data, there are signs that the module contributed to at least one goal—building students' social and political awareness. For example, in response to the post-module survey, a student from Eve's class wrote that the module "tied into the real-life world… and brought awareness to certain things," while another of her students wrote that the module "helped me understand both statistics and society in more in-depth ways." Other students discussed how they had heard about income inequality from social media or had observed social inequalities in their daily lives. Seeing statistics and larger-scale data on the topic, however, helped to make the phenomenon more real. One of Eve's students wrote in the post-module survey: "I thought that there wasn't as much of a difference in the wage gaps between males and females let alone between the different races," but then "it was enjoyable to calculate out the numbers to find out the honest truth behind income inequality." In a focus group interview, several students of different racial and ethnic backgrounds from Will's class shared similar reflections (with exchanges about media sources omitted):

133 Interviewer:	What did you guys know about income inequality before you jumped into these lessons?
134 Student 1:	To be honest, I thought it was a myth.
135 Interviewer:	You thought it was a myth?
136 Student 1:	Yeah.
137 Interviewer:	How about for others?

[….]

142 Student 2: I agree that I saw it in the media a lot, but I don't think we were ever presented with proper information behind it. Like you always see on the news, oh the wage gap between men and women, but there would never be any statistics behind it. So I think this is an interesting input to actually see the statistics behind it.

[...]

149 Interviewer: How about others… where have you heard about income inequality before?

[...]

153 Student 3: From like real life. Like I would obviously see that colored people are not making, or not just colored, but like Hispanics don't make as much as others.

154 Interviewer: How can you tell?

155 Student 3: Just because of the jobs that some are forced to do. Like you just know that they're not making as much as like a lawyer would make, and it's obvious. But if I …*[inaudible]* see a percentage or see something saying about income inequality, I'm just like okay, so it does exist. And then I see more and more, and it's obviously real.

Not only did analyzing data help to cement the reality of income inequality in some students' minds, but the scale of the inequalities helped to awaken awareness of the depth of the issue. One of Will's students wrote in the post-module survey: "The overall idea of having attributes to really compare [income] data from 2017 is what astonished me, I didn't realize so many people were making $10,000 to nothing." In her debriefing interview, Bella observed the same reaction among her students: "They really latched onto the fact that 25 percent of the country makes less than $15,000 a year and what that means, and really how shocking that is." The ability to explore multiple explanatory variables also helped to heighten students' sense of the persistence and deep-rooted nature of the wage gap between male and female employees. Rachel described her students' reactions to the last lesson of the module: "They got more and more outraged that no matter what the education level was, the male median [income] was always larger than the female median and sometimes quite a bit larger. And they just kept going, 'But that's not fair.'"

There are also signs that for some students, the module may have helped to instill a greater sense of agency in their learning. One of Will's focus group students said: "I thought [the module] treated us more like students and less like children… because it asked for our input as opposed to telling us, hey, this is [it], learn about it." Another of his students wrote in the post-module survey that "I liked the income inequality data more than the statistics lessons we usually do because of the fact that it felt like we were our own statisticians analyzing data." Although the module did not include activities specifically designed to promote students' collective agency, at least one of Julie's students suggested that seeing evidence of income inequality in the data helped to spark such agency. In a focus group interview, she said, "You always hear about it, like there's income inequality between different genders, and when we actually did the data it was like, oh wow, this is real. People aren't just making things up. This is a real problem, and maybe *hopefully we can figure something out"* (emphasis added).

At the same time, some challenges may have hindered critical data literacy or social justice learning outcomes. One challenge involved feelings of disempowerment among some students when studying income inequality. In her debriefing interview, Bella shared: "I was a little worried during the project that there was this sense of this is inevitable." Although she wanted to address this sentiment, she said that "I didn't really have any time to talk about what's being done… [like] pay equity, laws, and activism." Sasha shared similar reflections in her interview. She had noticed a "fatalism" among some students, which data on income inequality had reinforced. She voiced what she believed were students' thoughts: "Okay, now I can see that this is how it is for all Black men. So why do I bother? All the… crashing, crushing sense of inevitability, and I'm never going to be able to get out of the system." When asked how she handled these sentiments, Sasha said that she raised examples of social activism and protest movements that have made a difference in U.S. history, but she did not have time to go in depth with these topics.

7. Discussion

Prior research efforts to promote students' critical data literacy or to teach mathematics for social justice have typically involved smaller-scale classroom case studies and interventions in which developers have played an active role in facilitating learning activities (e.g., Brantlinger, 2013; Gutstein, 2003; Gutstein, 2006; Kahn, 2020; Rubel, 2016; Van Wart et al., 2020). This study adds to the literature by examining a critical data literacy intervention that was developed for implementation by high school mathematics teachers, without classroom facilitation support by project developers. Employing both quantitative measures and qualitative data, we explored changes in learning and affective outcomes among approximately 200 students of seven teachers of grade 12 non-AP mathematics classes in schools with high proportions of students from historically marginalized groups. We found that teachers were able to implement the module after receiving less than two days of professional development, and module use was associated with statistically significant growth in students' understanding of fundamental data literacy concepts and interests in statistics and data analysis. Our results suggest that the curriculum approach has potential to strengthen important *data literacy* outcomes—which we consider to be the foundation for critical data literacy—for diverse student populations. Study designs with randomized control group comparisons and strategies to reduce student attrition are needed, however, to establish the true efficacy of the module. Additional qualitative research would also help us understand why student outcomes may vary so widely across teachers and under what conditions the module may hold the greatest promise. Educators need such interventions to help all students of all backgrounds gain knowledge and practices that are essential for participation in today's data-driven society.

Several module components appeared to play large roles in supporting students' data literacy interests and understandings. One component involved the module's topical focus: income inequality. Consistent with theories of culturally relevant pedagogy (e.g., Aronson & Laughter, 2016; Ladson-Billings, 1995), students shared that the topic was compelling because it was a real-life issue relevant to their lives. What makes a topic or learning activity relevant, however, may have different interpretations (e.g., Ladson-Billings, 2014; Paris, 2012). Enyedy and Mukhopadhyay (2007) found that learning activities may be socially or culturally relevant in at least three ways: by drawing on contexts from students' daily lives, by attending to topics that students find valuable in their lives outside of school, and by adopting learning processes and norms that are familiar to students. The students in our study suggested additional nuances and details to the concept of relevance. A topic like income inequality may be relevant because it is an authentic problem that students feel they may face in real life—not an artificial question such as the relationship between GPA and height. It may be relevant by providing students with practical information to guide their own future life decisions, such as where to live or what occupations to consider for maximizing future earnings. Relevant topics may also involve social injustices that arouse one's moral indignation. For some students in our study, the topic of income inequality was relevant because it served individualistic or instrumental goals; for others, relevance connected to larger social, emotional, or moral concerns. Lee et al. (2021) recommend that learning designers seek data sources and topics with which students can connect at personal, cultural, or social and political levels. Students in our study appeared to find the topic of income inequality meaningful at all three of these levels.

Dynamic and interactive data features within CODAP also appeared to help engage students in data practices and build their understandings of important data concepts. In fall 2019 and early 2020, prior to the onset of the COVID-19 pandemic and during the period of this study, students and teachers said that statistics learning typically involved teacher lectures, independent problem-solving with textbook-based problems, and computations performed with graphing calculators. The prevalence of more didactic pedagogy and older forms of technology is in keeping with research that has documented the slow pace of reform in statistics education (Zieffler et al., 2018). The teachers who chose to participate in our study were attracted to the inquiry-based approach of our project model and may have needed the curriculum and professional development supports that we offered to move toward more student-centered forms of instruction. In these contexts, many students found it novel and exciting to investigate official and large-scale data from the U.S. Census Bureau using CODAP's interactive computer-based data visualization tools. Both students and teachers suggested that the ability to display, explore, and even manipulate the shapes and statistical measures of large data distributions—particularly the provocative and highly skewed distribution of individual incomes in the United States—helped students gain a deeper understanding of basic statistical concepts, such as when and why the mean and median may differ. Traditional methods of instruction typically fail to support such understandings (Konold & Higgins, 2003). Students showed the greatest growth in their understanding of measures of center based on their pre- and post-module assessment scores, suggesting that CODAP tools and module activities may have been especially helpful for learning in this domain. In addition, the Microdata Portal enabled students to draw and investigate repeated samples of varying sizes from available decennial census and ACS data. At least one teacher felt that this activity helped students develop an intuition for the level of accuracy associated with statistical estimates from samples of different sizes. This is an intuition that is challenging and complex to build (Chance et al., 2004; Garfield et al., 2015).

The use of person-level microdata also appeared to enhance students' data interests and learning. Public data that non-experts can access and analyze from sources like the U.S. Census Bureau typically provide information about people and their characteristics at aggregate levels, involving units of analysis such as states, metropolitan areas, counties, or neighborhoods. Although neighborhood-level or other geographic-level data can provide rich opportunities to examine spatial patterns in aggregate outcomes of interest (e.g., Kahn, 2020; Radinsky, 2014; Rubel et al., 2016; Van Wart et al., 2020), units of geography obscure the individuals within them and may make it more difficult for students to relate to the data. Students and teachers in our study indicated that by working with person-level data in CODAP, students could see and identify with the individual people in their data samples. By engaging in activities that focused attention on the original ACS questionnaire and on CODAP's linked displays of individuals in table rows and as dots in dot plots, at least some students seemed to develop a stronger understanding of the provenance of the data and the multiple ways in which data can be represented. Prior research points to the barriers that learning designers may face when trying to help students connect with secondary data that they have not collected themselves (e.g., Lee & Wilkerson, 2018). Learning activities involving person-level microdata, when available in student-friendly formats, may help remove some of these barriers.

The tools and activities that we employed to advance multivariable thinking—an important aspect of data literacy—may have had an impact within our study sample, given positive findings from our pre- and post-module assessment data. The opportunity to examine relationships among multiple variables also appeared to support aspects of *critical data literacy* in at least some of our participants, as suggested by our qualitative data. Some students indicated that they developed greater awareness of income inequality between higher- and lower-income earners by seeing the large numbers of individuals earning low wages, the low numbers earning high wages, and how this pattern has persisted (and worsened) over multiple decades. Uncovering large-scale and persistent data patterns helped to validate some students' experiences, showing them that observations of income disparities from their own lives were not flukes but were systemic in society. Students who had been previously skeptical of claims about income inequality also found the data patterns credible and convincing, perhaps because they had conducted in-depth activities considering the origins and limitations of their data as well as the levels of accuracy afforded by their data samples.

Students also became much more convinced that sex discrimination may exist by examining differences in the typical wages of male and female employees and finding that the wage gap does not disappear (and sometimes grows) after they had the opportunity to control for multiple other variables. Enyedy and Mukhopadhyay (2007) raised warnings that in certain contexts, students may not examine data critically and may simply draw conclusions that accord with their own preconceptions. Our work suggests that one way to guard against this outcome is to design activities in which students investigate claims about social disparities—just as scholars of TMSJ advocate—with tools and supports that encourage students to explore multiple competing explanations with multivariable data. Such activities have strong potential to advance aspects of critical data literacy that align with two of Gutstein's (2003, 2006) goals for social justice pedagogy: building students' awareness of larger social and political forces in their lives and fostering their agency to advocate for social change with data.

Designing and implementing such learning activities in classrooms with diverse populations will face challenges. Philip et al. (2016) note how discussions of social inequality in the United States almost inevitably raise issues of race and how conversations about race without skilled facilitators can be harmful to students. Although we did not observe negative exchanges involving race in our participating classrooms, and although teachers and students did not report any such exchanges, we cannot be sure that our classrooms avoided them entirely. Educators who wish to promote critical data literacy among students need to know not only how to facilitate students' learning of technical data literacy skills, but also how to guide students through conversations that explore why and how different groups may experience unequal social, economic, or political outcomes. In addition, there is a risk that extended investigations of topics that highlight systemic social injustices can feel demoralizing for students, especially for those students from historically marginalized groups. As suggested by some of our participating teachers, a strategy for countering such responses may be to include examples of individual and collective actions that have led to positive social change. Within the microcosm of classrooms, teachers can also work to establish norms and facilitate learning activities that value and build upon the contributions of students from all backgrounds to help students experience and envision more equitable social interactions and arrangements (Alim & Paris, 2017). Educators will need time, materials, and support to pursue these strategies, however, if efforts to promote critical data literacy are to reach their full potential.

Acknowledgement

This work was supported by the National Science Foundation (NSF) under Grant No. DRL 1813956. Opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the NSF.

References

Alim, H. S., & Paris, D. (2017). What is culturally sustaining pedagogy and why does it matter? In D. Paris & H. S. Alim (Eds.), *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world* (pp. 1–24). Teachers College Press.

Aronson, B., & Laughter, J. (2016). The Theory and practice of culturally relevant education: A Synthesis of research across content areas. *Review of Educational Research, 86*(1), 163–206. https://doi.org/10.3102/0034654315582066

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II)* (2nd ed.). American Statistical Association.

Baumer, B. S., Garcia, R. L., Kim, A. Y., Kinnaird, K. M., & Ott, M. Q. (2022). Integrating data science ethics into an undergraduate major: A Case study. *Journal of Statistics and Data Science Education, 30*(1),15–28. https://doi.org/10.1080/26939169.2022.2038041

Berry, R. Q., Conway, B. M., Lawler, B. R., & Staley, J. W. (2020). *High school mathematics lessons to explore, understand, and respond to social injustice*. Corwin Press.

Bhargava, R., Deahl, E., Letouze, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data.* Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews.

Brantlinger, A. (2013). Between politics and equations: Teaching critical mathematics in a remedial secondary classroom. *American Educational Research Journal, 50*(5), 1050–1080. https://doi.org/10.3102/0002831213487195

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The Role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1). https://doi.org/10.5070/T511000026

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning, and thinking* (pp. 295–323). Springer.

Cuoco, A., Goldenberg, E. P., & Mark, J. (1996). Habits of mind: An Organizing principle for mathematics curriculum. *Journal of Mathematical Behavior*, *15*(4), 375–402. https://doi.org/10.1016/S0732-3123(96)90023-1

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Dee, T., & Penner, E. (2016). *The Causal effects of cultural relevance: Evidence from an ethnic studies curriculum* (CEPA Working Paper No.16-01). Stanford Center for Education Policy Analysis. http://cepa.stanford.edu/wp16-01

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28–58. https://doi.org/10.52041/serj.v6i2.483

Delpit, L. D. (1988). The Silenced dialogue: Power and pedagogy in educating other people's children. *Harvard Educational Review, 53*(3), 280–298. https://doi.org/10.17763/haer.58.3.c43481778r528qw4

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Sage Publications.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics, 26*(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

Engel, J. (2016). Open data at the interface of mathematics and civics education: Challenges of the data revolution for the statistics curriculum. *Journal of Mathematics and Statistical Science*, *2*(5), 264–273.

Enyedy, N., & Mukhopadhyay, S. (2007). They don't show nothing I didn't know: Emergent tensions between culturally relevant pedagogy and mathematics pedagogy. *Journal of the Learning Sciences, 16*(2), 139–174. https://doi.org/10.1080/10508400701193671

Finzer, W. (2013). The Data science education dilemma. *Technology Innovations in Statistics Education, 7*(2), 1–9. https://doi.org/10.5070/T572013891

Frankenstein, M. (2009). Developing a criticalmathematical numeracy through *real* real-life word problems. In *Words and Worlds: Modelling Verbal Descriptions of Situations* (pp. 111–130). Sense Publishers.

Frankenstein, M. (2013). Reading the world with math. In E. Gutstein & B. Peterson (Eds.), *Rethinking mathematics: Teaching social justice by the numbers* (pp. 30–39). Rethinking Schools.

Freire, P. (1970). *Pedagogy of the oppressed* (30th ed.). Continuum International.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 7*(1), 1–51. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Garfield, J., delMas, B., Chance, B., & Ooms, A. (2006). *Summary statistics for a national sample of undergraduates: Comprehensive assessment of outcomes for a first course in statistics (CAOS).* The Web ARTIST Project.

Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics, 88,* 327–342. https://doi.org/10.1007/s10649-014-9541-7

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53,* (2), 106–116. https://doi.org/10.1177/0022487102053002003

Gay, G. (2013). Teaching to and through cultural diversity. *Curriculum Inquiry, 43*(1), 48–70. https://doi.org/10.1111/curi.12002

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal, 16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education, 34*(1), 37–73. https://doi.org/10.2307/30034699

Gutstein, E. (2006). *Reading and writing the world with mathematics.* Routledge.

Gutstein, E. (2013). Home buying while Brown or Black. In E. Gutstein & B. Peterson (Eds.), *Rethinking mathematics: Teaching social justice by the numbers* (pp. 61–66). Rethinking Schools.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016). *The Age of analytics: Competing in a data-driven world.* McKinsey Global Institute.

Jacobbe, T., Case, C., Whitaker, D., & Foti, S. (2014). Establishing the validity of the LOCUS assessments through an evidenced-centered design approach. In *Sustainability in statistics education. Proceedings of the ninth international conference on teaching statistics (ICOTS9).* International Statistical Institute. https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7C2_JACOBBE.pdf?1405041692

Kahn, J. (2020). Learning at the intersection of self and society: The Family geobiography as a context for data science education. *Journal of the Learning Sciences, 29*(1), 57–80. https://doi.org/10.1080/10508406.2019.1693377

Kokka, K. (2020). Social justice pedagogy for whom? Developing privileged students' critical mathematics consciousness. *The Urban Review, 52*(2), 778–803. https://doi.org/10.1007/s11256-020-00578-8

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). National Council of Teachers of Mathematics.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: A.k.a. the remix. *Harvard Educational Review, 84*(1), 74–84. https://doi.org/10.17763/haer.84.1.p2rj131485484751

Lee, V. R., & Wilkerson, M. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects.* National Academies of Sciences, Engineering, and Medicine. https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_189500.pdf

Lee, V.R., Wilkerson, M., & Lanouette, K. (2021). A Call for a humanistic stance toward K-12 data science education. *Educational Researcher, 50*(9), 664-672. https://doi.org/10.3102/0013189X211048810

Levels of conceptual Understanding in Statistics [LOCUS] (2002). *Question browser.* https://locus.statisticseducation.org/professional-development/questions/by-grade/high-school-%28grades-9-12%29?page=&type=prodev_const_response_question

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement, 70*, 647–671. https://doi.org/10.1177/0013164409355699

Louie, J. (2022). *Critical data literacy: Creating a more just world with data.* National Academies of Sciences, Engineering, and Medicine. https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D16254F310D01BBDA873920E4EFB8151F2D8334181AA

Louie, J., Roy, S., Chance, B., Stiles, J., & Fagan, E. (2021). Promoting interest and skills in statistical and multivariable thinking with social justice data investigations. In R. Helenius & E. Falck (Eds.), *Statistics education in the era of data science: Proceedings of the International Association for Statistical Education (IASE) 2021 Satellite Conference*. https://iase-web.org/documents/papers/sat2021/IASE2021%20Satellite%20133_LOUIE.pdf?1649974212

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The Next frontier for innovation, competition, and productivity*. McKinsey Global Institute. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press. https://doi.org/10.17226/25104

National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA Center & CCSSO]. (2010). *Common core state standards for mathematics*.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A Critical literacies approach to enhancing understandings of personal digital data. *New Media and Society, 21*(2), 419437. https://doi.org/10.1177/1461444818799523

Paris, D. (2012). Culturally sustaining pedagogy: A Needed change in stance, terminology, and practice. *Educational Researcher, 41*(3), 93–97. https://doi.org/10.3102/0013189X12441244

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction, 34*(4), 361–388. https://doi.org/10.1080/07370008.2016.1210418

Radinsky, J., Hospelhorn, E., Melendez, J. W., Riel, J., & Washington, S. (2014). Teaching American migrations with GOS census webmaps: A Modified "backwards design" approach in middle-school and college classrooms. *Journal of Social Studies Research, 38, 143-158.* https://doi.org/10.1016/j.jssr.2014.02.002

Raffaghelli, E. J. (2020). Is data literacy a catalyst of social justice? A Response from nine data literacy initiatives in higher education. *Education Sciences, 10*(9), 233. https://doi.org/10.3390/educsci10090233

Ridgway, J. (2015). Implications of the data revolution for statistics education: The Data revolution and statistics education. *International Statistical Review, 84*(3), 528–549. https://doi.org/10.1111/insr.12110

Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016). Teaching mathematics for spatial justice: An Investigation of the lottery. *Cognition and Instruction, 34*(1), 1–26. https://doi.org/10.1080/07370008.2015.1118691

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences, 29*(1), 154–164. https://doi.org/10.1080/10508406.2019.1705665

Sproesser, U., Engel, J., & Kuntze, S. (2016). Fostering self-concept and interest for statistics through specific learning environments. *Statistics Education Research Journal, 15*(1), 28–54. https://doi.org/10.52041/serj.v15i1.256

Van Wart, S., Lanouette, K., & Parikh, T. (2020). Scripts and counterscripts in community-based data science: Participatory digital mapping and the pursuit of a third space. *Journal of the Learning Sciences, 29*(1), 127–153. https://doi.org/10.1080/10508406.2019.1693378

Weiland, T. (2017). Problematizing statistical literacy: An Intersection of critical and statistical literacies. *Educational Studies in Mathematics, 96*, 33–47. https://doi.org/10.1007/s10649-017-9764-5

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3275

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education*. *Springer International Handbooks of Education* (pp. 37–70). Springer. https://doi.org/10.1007/978-3-319-66195-7_2

Appendix 1

Sample Items from the Pre- and Post-Module Data Literacy Assessment

I. Sampling and Data Collection

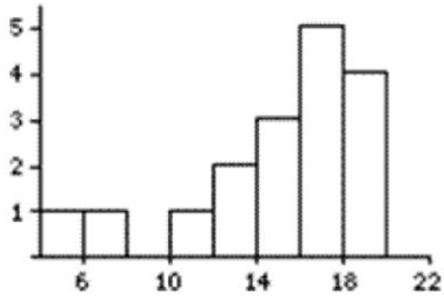
Nathan read that students tend to carry backpacks that are very heavy. He decides to collect data to determine if this is true for his school. Which of the following would be the most appropriate data for Nathan to collect?

a. The number of books in students' backpacks
b. Whether or not students carry a backpack to school
c. The weight of the backpacks that students carry to school
d. The number of days per week students carry a backpack to school

Source: Adapted from Levels of Conceptual Understanding in Statistics (LOCUS), 2022.

II. Measures of Center

In the graph of Quiz Scores below, which estimates of the mean and median are most likely to be correct?



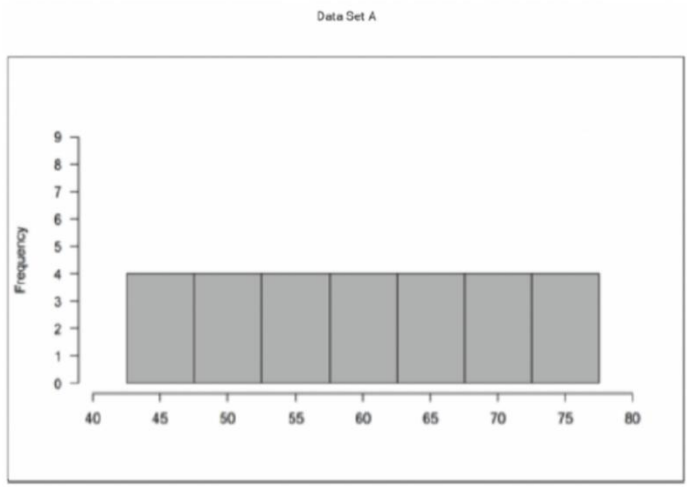
Quiz Scores

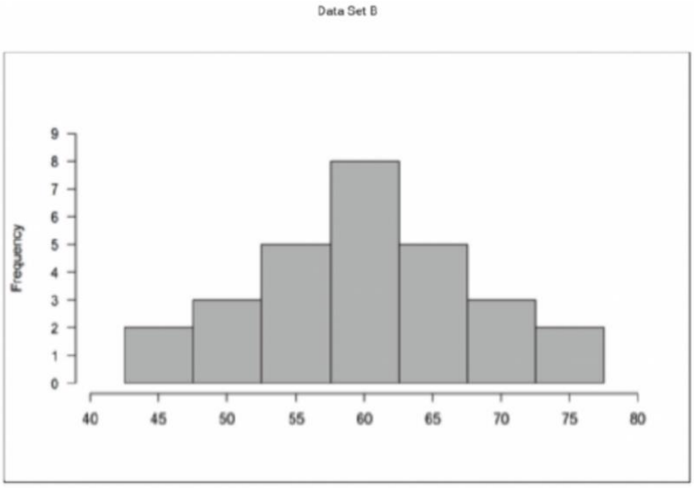
a. median = 13.0 and mean = 12.0
b. median = 14.0 and mean = 15.0
c. median = 16.0 and mean = 14.3
d. median = 16.5 and mean = 16.2

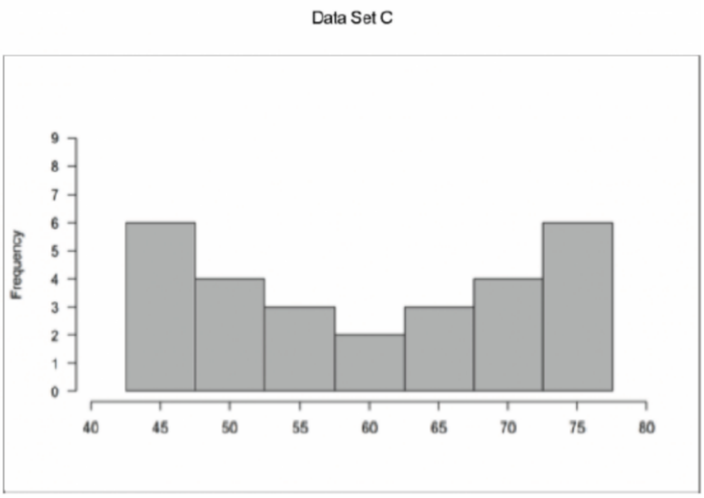
Source: Adapted from del Mas et al. (2007).

III. Measures of Variability

Three data sets are summarized in the histograms below.







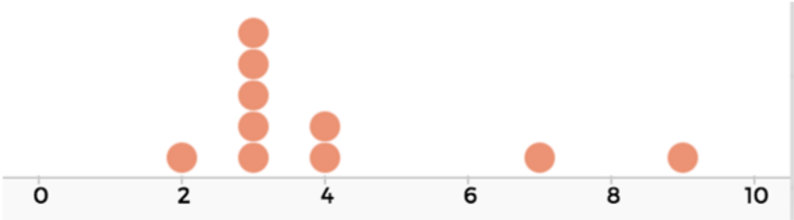
Which set of data varies the most from its mean?

a. Data set A
b. Data set B
c. Data set C
d. The variability from the mean is the same for all three data sets.

Source: LOCUS (2022).

IV. Data Representation

The dotplot below shows the quiz scores for each student who took a quiz. The quiz had 10 questions.

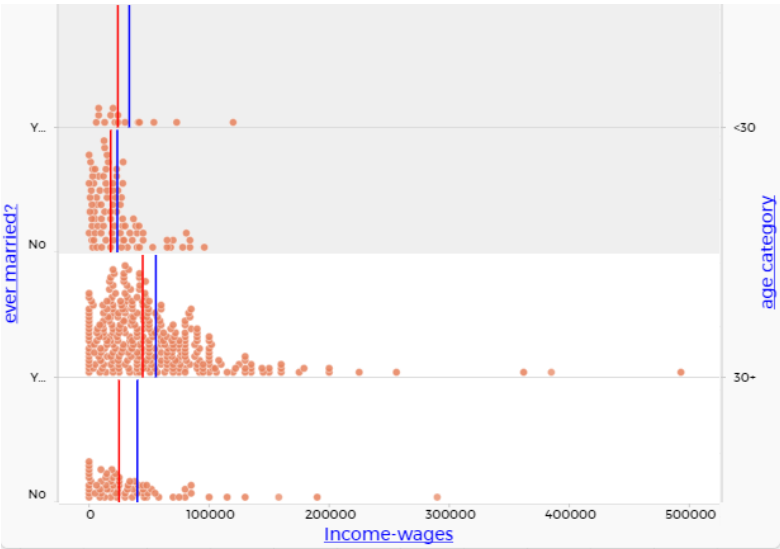


How many students received scores higher than 4?
a. 1
b. 2
c. 3
d. 4

Source: Authors.

V. Multivariable Thinking/Reasoning

Below is a set of graphs that add data on a third variable: age. The same 476 employed individuals in the graph above are now split into two age categories: whether they are under 30 years old (<30), or 30 years old or above (30+).



Source: Authors.

Based on the additional information in the graph above, which of the following is an appropriate statement?

a. There is no difference in income-wages for people who have been married compared to people who have never been married once you control for a person's age category.

b. There is no difference in income-wages for people under age 30 compared to people age 30+ once you control for whether they have ever been married.

c. People who have been married earn higher income-wages, on average, than people who have never been married, even after controlling for their age category.

d. People age 30+ earn lower income-wages, on average, than people under age 30, even after controlling for whether or not they have ever been married.

Appendix 2

Table 7. Example interest survey items and associated subscales

Item	Scale
Our statistics lessons over the past couple of weeks sparked my interest in investigating data.	Situational interest
What we learned in *statistics class over the past couple of weeks* is important to me.	
Investigating data is one of my favorite activities.	Individual interest
I like to think about statistics and data even when I'm outside of my statistics class.	
Understanding tasks with diagrams and statistical data is easy for me.	Self-concept
I am good at solving statistical problems.	
It is important to me to be a person who can analyze data statistically.	Perceptions of value
Statistics and working with data are useful for me to know.	

Note. Phrases in italics were replaced with "the Income Inequality data lessons" in the post-module survey. Each item had a seven-point rating scale, ranging from 1 (Strongly disagree) to 7 (Strongly agree).