International Society of the Learning Sciences

Aesthetics of Authenticity for Teachers' Data Set Preferences

Victor R. Lee, Stanford University, vrlee@stanford.edu
Victoria Delaney, Stanford University, vldocherty@stanford.edu

Abstract: This paper explores secondary school teachers' aesthetic judgments of data sets for prospective use to teach data science. Twelve teachers were interviewed and asked to examine and select from side-by-side data sets the one that would provide a more authentic experience for students. Situating our work in the "Knowledge in Pieces" epistemology, we drew upon diSessa's (1993) notion of an aesthetic – a knowledge system for appraisal and judgment. We examined what sensitivities were expressed and how judgments were made and supported. We observed that teachers drew upon at least three senses of authenticity to characterize and select data sets: authentic as messy, authentic as requiring more work, and authentic as involving computation. Identification of these ways of determining authenticity represents an initial step in better understanding how teachers appraise data that could be used in their teaching.

Introduction

In their review of how data are being used as part of investigations and inquiry experiences in middle and secondary classrooms, Lee & Wilkerson (2018) examined the current state of learning sciences research related to the use of curated and public data sets in classrooms. Specifically, they found substantially less guidance for how educators can best use these kinds of data, and they stressed that this was an area where more research and support was needed. As researchers in this space who expect that the design of rich data experiences will become a more persistent concern for the learning sciences, we are trying to address these gaps. We are imagining a future where teachers are able to obtain and introduce quantitative data sets into their classrooms much in the same way that a teacher might select and include what they consider to be a related or useful news article or website that they have discovered online as an instructional resource. In that scenario, some sets of data will be selected and be judged to be more or less useful in their teaching. We ask in this paper how those judgments get made.

Beyond being a relatively understudied area in the emerging space of pre-collegiate "data science education", we subscribe to a model of classroom instruction that involves teachers designing their own or adapting existing curricula (Brown & Edelson, 2003). This makes teacher selection of data sets one of the many design decisions that get made. However, in thinking about this design decision, it is unclear the process that teachers will go through in their decision making. We also have little existing research on which to base teacher learning experiences that will help teachers to be even more thoughtful and reflective of their data selection decisions. Rather, research on teachers and their thinking related to data sets focuses largely on their conceptual understandings (often presented as misunderstandings) of certain core topics involved in data analysis (Batanero, Burrill, & Reading, 2011).

As there are now growing calls to develop an area of data science education (Lee & Wilkerson, 2018) and for learning scientists to contribute to research and design supportive of it (Wilkerson & Polman, 2020), we believe some new work on teacher thinking about data sets is needed. To motivate this in terms that are more practice-oriented, we might imagine a situation where a high school teacher wants to bring in a real data set from professional research that relates to natural disasters. If they live in an area that is known to have major fault lines and seismic activity, then seismology data might be more desirable. However, if the teacher is working in an area where hurricanes and flooding are the major concern, then data on hurricanes might be more desirable. Decision making could be as simple as deciding what is more locally relatable, but there could still be a number of potential complicating factors. The quality or amount of data may differ, or the structuring of the data – such as using numbers on a logarithmic scale - may introduce challenges that lead teachers to be more reluctant to use them. Furthermore, a geography teacher may have different priorities than an Earth Science teacher, and they may have different priorities still from a mathematics teacher.

Regardless of whether teachers are doing the work of selecting data sets or whether curriculum or software designers have pre-selected data for them, some consideration of how teachers judge and appreciate data is needed. In this paper, we report on some initial efforts to examine that. Our focus is on data that is thought to be supportive of an authentic learning experience related to data science. The theoretical perspective we draw from is based in the "knowledge in pieces" epistemology (diSessa, 1988). In the section below, we discuss theoretical bases involving aesthetics and authenticity. Following that, we present some initial results in our efforts to understand teachers' aesthetics for authenticity.

Theoretical Perspectives

Knowledge and aesthetics

This work is situated within what learning scientists refer to as the "Knowledge in Pieces" (KiP) epistemology (diSessa, 1988), which is most prominent in conceptual change research. It has also been recently used to model knowledge in interaction (diSessa, Levin, & Brown, 2016). As a theoretical perspective, KiP provides an alternative to descriptions of student knowledge that are critiqued for using imprecise, or arguably inaccurate, terminology (diSessa, Gillespie, & Esterly, 2004). A common use of KiP has been to challenge cognitive models of knowledge that make strong claims about context invariance. For instance, the claim that students come to instruction with rigid misconceptions that must be disproven is one that has come under critique from the KiP perspective. The proposed alternative to a misconceptions view is to think of student knowledge as involving more atomistic pieces of knowledge that are cued in response to contextual information (Smith, diSessa, & Roschelle, 1993). One interaction in which this has been made more apparent is in Piagetian-style clinical interviews. In such interviews, occasions where students seemingly contradict themselves or change their minds within minutes of having already given a seemingly confident and definitive explanation for a phenomenon are often cited as a demonstration of how a KiP perspective can be especially useful (Sherin, Krakowski, & Lee, 2012).

Some of the most prominent KiP work emphasizes the specification of certain kinds of cognitive resources that play key roles in dynamic reasoning processes (e.g., Hammer et al., 2005). Another related area of work within the KiP tradition is the identification and description of what diSessa calls *aesthetics* (diSessa, 1993). Aesthetics are "rich but structurally limited knowledge systems, which, notwithstanding their richness, appear fluid, data driven, and involve situation-specific reasoning…and idiosyncratic justification." (p. 187). As diSessa describes them, aesthetics provide judgments of similarity and preference. They also serve to provide hunches and intuitions with varying levels of confidence. Aesthetics are involved in a range of situations. One of the most well-known is the sense of mechanism that is invoked in commonsense physics reasoning. For instance, the sense of mechanism for physics helps students to predict behaviors of objects in motion or determine where forces are involved in a physical situation (diSessa & Sherin, 1998). Another aesthetic is the knowledge system that develops in order to recognize the law of large numbers within different probabilistic situations (Wagner, 2006). While these are mathematics and science-oriented examples, diSessa (1993) is explicit about stating that aesthetics are applicable to a range of domains. Knowledge about interpersonal relations is suspected to part of a social aesthetic. Indeed, this suspicion is explored and applied to a case of a teacher's ideological reasoning by Philip (2011). In that example, the aesthetic system serves to make judgments about students and explain why some students succeed or fail at being students.

One goal of the current paper is to provide some initial description of aesthetics involved in evaluating data sets. In particular, the aesthetics that are being considered are those of high school mathematics and statistics teachers. We selected these teachers because they were likely candidates for teaching data science as a standalone topic or course. While we focus on mathematics and statistics teachers, we do note that where and how data science should or eventually will be integrated into the curriculum is unsettled. Informally speaking, based on research workshops and summits, interest has been expressed from those in the computer science education, math & statistics education, and science education communities. Regardless, there are cases of standalone data science courses being deployed in large school systems (see Lee & Delaney, 2021) and being deemed eligible because they count as fulfilling the same requirements as that of a high school statistics course.

Authenticity

Learning sciences has often stressed authenticity as a valued priority in the design of learning experiences. Accessible, but authentic forms of scientific inquiry, are often pursued (Edelson & Reiser, 2006). This may involve supporting students in scientific modeling (Schwarz et al., 2009) or providing students access to scaffolded versions of professional tools like GIS systems (Edelson, 2004). In this treatment of authenticity, similarity or congruence with respect to professional practice are prioritized.

At the same time, research in the learning sciences has also treated community-based forms of knowing, understanding, and acting in the world as authentic, although those instantiations may not bear immediate resemblance to professional disciplinary practice. A well-known example comes from Lave and colleagues who documented how learning how to do solve mathematics problems to determine the best buy of a product in a traditional classroom can look quite different from what actually happens when one is determining best buy of a product while shopping in a real-world setting (Lave, Murtaugh, & de la Rocha, 1984).

That we can apply authenticity to these different situations suggests that there may be different frames of reference for what makes something authentic. Indeed, there is evidence for this in some of the psychology literature. In his summary of prior research on authenticity, Newman (2019) identified multiple ways by which we judge authenticity, at least with respect to objects. These include historical, categorical, and values authenticity. Historical authenticity implies a connection to a particular person, time, or place. An authentic Picasso painting would be one that was actually painted by Pablo Picasso. Categorical authenticity would match existing beliefs and requirements for a category to which the object belongs. Authentic Neopolitan pizza could be purchased and enjoyed outside of Naples provided that it was made with the correct ingredients and followed a specific preparation process. Values authenticity would involve consistency between internal state and external expression. While two politicians would say that they support racial justice, one might come under fire for being inauthentic whereas another might be seen as actually authentic.

It is unlikely that these treatments of authenticity are exhaustive, especially considering these were presented primarily with respect to objects. However, this illustrates that authenticity can be complex and subject to various evaluative criteria. In light of that, the question we examined was what data sets would our participating high school teachers deem as providing a more authentic experience for students. Our goal was to get some traction on understanding an authenticity aesthetic, as it relates to teaching high school data science.

Methods

Research Participants

The participants in this study were 12 secondary school math and statistics teachers. These teachers were not recruited through a randomized process. Rather, they were solicited from existing contacts in schools and through snowball sampling. There were 6 men and 6 women who participated. These individuals had a range of 2 to 31 years of teaching experience, with the average being 13 years. These teachers came from a mix of public, private, and charter schools in the US, and some taught other courses, such as computer science, as well.

Procedure

The teachers participated in a videorecorded interview that lasted approximately one hour. These interviews were completed during the COVID-19 pandemic and thus were done via video communications (i.e., Zoom). The bulk of the interview involved the teachers being presented with two pairs of data sets as tables in CODAP, a free web-based educational tool for examining and visualizing data. During the interview, the teachers were provided a brief introduction to CODAP and shown what were its visualization capabilities using a data set that was provided in the software. Because these were done via Zoom, the interviewer explained that she would show the teachers the data sets as tables and at any time they could request that the view of the interviewer's shared screen be changed (e.g., scrolling) or that a data visualization be made. Surprisingly, very few teachers asked for visualizations to be made.

One pair of presented data sets were of passengers on the Titanic. Titanic data were selected as they are common in online and undergraduate data science courses and high school statistics courses. One Titanic data set ("Titanic A") came from the data website, Kaggle.com, which is used by data science enthusiasts, holds data sets on a wide range of topics, and hosts competitions for users. The other Titanic data set ("Titanic B") came from the *Introduction to Data Science* curriculum (Gould et al., 2018) used in a number of high schools in the United States. Titanic A had 891 cases, whereas Titanic B had 1000 cases.

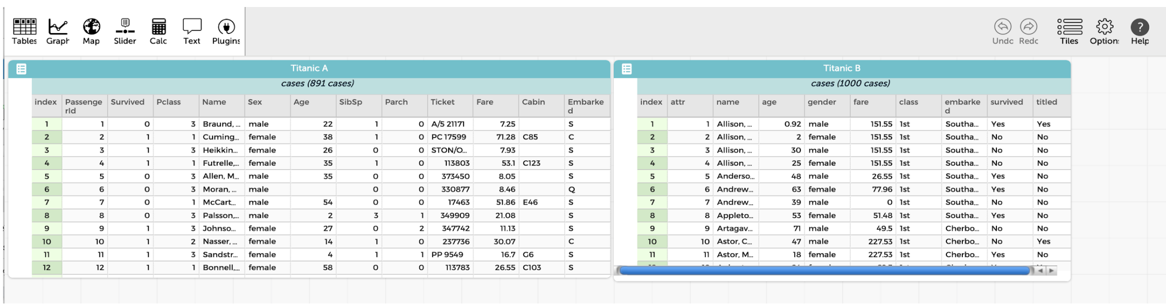
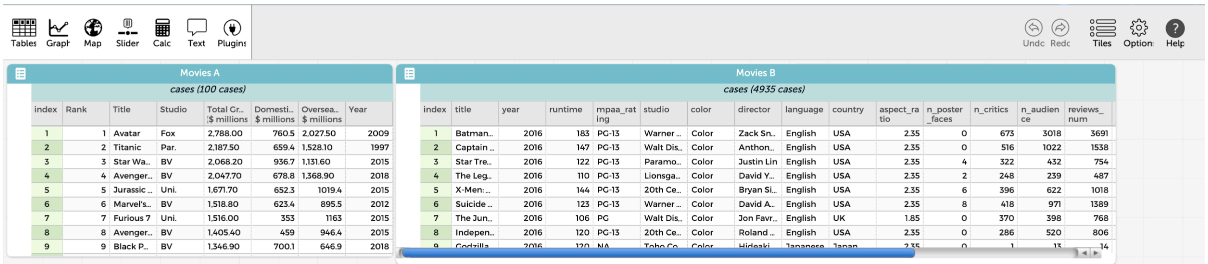


Figure 1. The Titanic A and Titanic B data sets.

The other pair were of popular movies and their box office revenue. One came from the *Bootstrap:Data Science* curriculum ("Movies A"). *Bootstrap:Data Science* is one of multiple *Bootstrap* curricular units that were developed to integrate computer science into curricula. The other movies data set ("Movies B") came from the above mentioned *IDS* curriculum. Movies B was much larger than Movies A in that it had 4935 cases while Movies A had 100. These data sets were downloaded unaltered from their source curricula resource sites. While there was over an order of magnitude difference in the sizes of the two data sets, we chose to use them as they were because that was how they were made available in their respective curricula.



index	Rank	Title	Studio	Total Gr... ($ millions	Domesti... $ millions	Oversea... $ millions	Year
1	1	Avatar	Fox	2,788.00	760.5	2,027.50	2009
2	2	Titanic	Par.	2,187.50	659.4	1,528.10	1997
3	3	Star Wa...	BV	2,068.20	936.7	1,131.60	2015
4	4	Avenger...	BV	2,047.70	678.8	1,368.90	2018
5	5	Jurassic ...	Uni.	1,671.70	652.3	1019.4	2015
6	6	Marvel's...	BV	1,518.80	623.4	895.5	2012
7	7	Furious 7	Uni.	1,516.00	353	1163	2015
8	8	Avenger...	BV	1,405.40	459	946.4	2015
9	9	Black P...	BV	1,346.90	700.1	646.9	2018

index	title	year	runtime	mpaa_rating	studio	color	director	language	country	aspect_ratio	n_poster_faces	n_critics	n_audience	reviews_num
1	Batman...	2016	183	PG-13	Warner ...	Color	Zack Sn...	English	USA	2.35	0	673	3018	3691
2	Captain ...	2016	147	PG-13	Walt Dis...	Color	Anthon...	English	USA	2.35	0	516	1022	1538
3	Star Tre...	2016	122	PG-13	Paramo...	Color	Justin Lin	English	USA	2.35	4	322	432	754
4	The Leg...	2016	110	PG-13	Lionsga...	Color	David Y...	English	USA	2.35	2	248	239	487
5	X-Men: ...	2016	144	PG-13	20th Ce...	Color	Bryan Si...	English	USA	2.35	6	396	622	1018
6	Suicide ...	2016	123	PG-13	Warner ...	Color	David A...	English	USA	2.35	8	418	971	1389
7	The Jun...	2016	106	PG	Walt Dis...	Color	Jon Favr...	English	UK	1.85	0	370	398	768
8	Indepen...	2016	120	PG-13	20th Ce...	Color	Roland ...	English	USA	2.35	0	286	520	806
9	Godzilla	2016	120	NA	Toho Co...	Color	Hideaki	Japanese	Japan	2.35	0	1	13	14

Figure 2. The Movies A and Movies B data sets.

We would also like to note that *B:DS* and *IDS* were selected because their development was funded by the United States' National Science Foundation, and they are among the only complete secondary data science curricula freely available online. In a separate paper (Lee & Delaney, 2021), we have reported on some content analyses of these two curricula and noted some differences between the two. Among those differences were in how large and how varied the data sets were used throughout each curriculum. Those differences motivated us to inquire about how teachers would judge these different data sets. At the same time, the data sets were different topically across the two units. Movies was one common topic, and Titanic was selected for the reasons stated above.

During the interview, the teachers were asked which data set will give students the most authentic experience of working with data in the context of a course they created that addressed data science. Those segments of the interview were our focus for this paper.

Analysis

The most common form of analysis for KiP oriented work is *knowledge analysis* (diSessa, Sherin, & Levin, 2016). Knowledge analysis typically involves video or audio capture of knowledge in use. For instance, interviews that involve participant sense-making about everyday activities in 'scientific' terms is one common interaction genre. Other approaches can involve paired problem-solving or work with a microworld.

Regardless of what type of interaction is recorded, those records are transcribed and then episodes are selected for iterative review and theory building. This may involve systematic and comprehensive coding, although when that is done, the goal of coding is to support theory building rather than emphasize frequency. Theory generation as part of knowledge analysis adheres to several core principles including a commitment to describing knowledge in terms of mental representations and fine grained and nuanced detail of those representations in action that is tightly connected to the presented data.

Reports of knowledge analysis tend to involve episodes of reasoning on the time scale of seconds to minutes with detailed description of knowledge elements at work. The reports are evaluated with respect to the adherence of theoretical claims about knowledge to the provided transcript excerpts and the potential for extending those claims to situations beyond those that are presented. Knowledge analysis reporting generally asks the reader to decide, assuming a KiP perspective, whether they are convinced of theoretical assertions that were motivated by the data.

For the current study, the interviews were transcribed and subject to multiple coding passes and joint interpretation sessions involving the authors. The coding passes varied from labeling content of what was discussed to what was being referenced within a data set to the sense of authenticity that was being evoked by the teacher. Some categorization of preferences was also done. Ten out of the twelve teachers picked Titanic A over Titanic B as providing a more authentic experience. For the Movies data sets, 11 picked Movies B.

Results

Authenticity as Messy

Part of the aesthetic appeal for some of the teachers appeared to be based on the sense of authenticity as the data being 'messy'. To illustrate, we provide an excerpt from an interview with Kiera. The interviewer was trying to get clarification from Kiera as to why she had just picked Titanic A over Titanic B. Kiera began by talking about Titanic B.

Kiera: I was saying that B has exactly 1000 cases, which is just such a nice number, you could make percents out of that easier, but is that just a coincidence, or is that manipulated data, whereas A seems a little bit more raw, I guess. B seems a little more cleaned up. So I guess if we're talking about the word authentic, I guess A.

While Kiera had many points to make, the one that is most relevant currently is in the second half of her utterance. She described Titanic A as "a little bit more raw". In contrast, Titanic B could be "manipulated data". If it was not manipulated, then for it to have 1000 cases would have to be a "coincidence" rather than to be expected. Moreover, the number 1000 had the quality of being "nice". Given just this statement, it would be hard to ascertain with confidence. It could be that it has something to do with 1000 being a decade number (i.e., multiple of 10), which has familiarity in the number of fingers we have on two hands (Domahs et al., 2015), our monetary system, and in how we often count and compute numbers in many languages (Murata, 2004). Visually, it also has repetition of three zeroes, making it feel less random as too much repetition is seen as less likely in chance events (Nickerson, 2002). Regardless, the aesthetic system extracts the number of cases and infers that it is special and less probable, making Titanic A more authentic. This observation appeared for multiple teachers in our study.

Authenticity as messy also appeared in other forms. Don, quoted below, was sensitive to the presence of organization in Titanic B but not in Titanic A.

Don: Yeah, so, I think that-they [students] would find it challenging that it's-well B is now organized alphabetically…If I wanted to be authentic, I think that I would probably choose A, because the data-is not organized, it's a little bit less clear on what's happening there, some of-like, the cabin is only listed for some of the people-Like, that's the same way, in the real world, you can't just get the 300 million people in the United States and-and you chose 1,000, you won't reach the thousand.

In this initial part of this selection, Don was conveying that his appraisal was that Titanic A was "not organized", or as we characterize it, disordered. Part of this disorganization came as a contrast from alphabetical order he recognized in Titanic B from looking at the *Name* column. From there, he continued with other ways he sees disorder – namely, the presence of empty cells appearing in the *Cabin* column ("the cabin is only listed for some of the people"). This suggests that for Don, detecting incompleteness is part of authenticity. To support that, Don made an analogy to "the real world" where if one were tasked with contacting 1000 people in the United States as part of a data collection effort, the thousand could not be reached. Authentic data experiences should involve encounters with incomplete data. Contacting 1000 people and succeeding in reaching them had a feeling of being orderly, whereas not getting all information was more disordered and to be expected.

Authenticity as Requiring More Work

While we saw Don state that parts of Titanic A were "a bit less clear" to refer to disorder or messiness, this appeared to also be connected to an aesthetic judgment related needing to do more work with the data in order to make inferences from it. There are indications of some underlying resistance or additional work. This was elaborated in the same utterance, of which another excerpt is provided below.

Don: So I feel like A is more challenging to look at, for, it's labeled a little bit less clearly, organized a little less, a little more abstract, *survived* is numbers and the *Pclass*-but, I think that then the flip side to that, I guess, is that B is a little bit easier for kids to deal with. And then, they could still-A requires the challenge of figuring out what the heck is going on, and then doing the data analysis. Whereas B is slightly, but not perfectly, organized, in a way that the kids would be able to attack the learning objectives of the class a little bit easier. But A is closer to what real life data would actually look like.

In some way, Titanic A required that students work more and work harder. Evidence for this appeared with his comparative statement that "B is a little easier for kids to deal with" and "A requires the challenge of figuring out what the heck is going on". Part of the work to be done comes from understanding the header labels, as they had, such as *PClass*. (At a separate time, Don stated "*Pclass*, I'm guessing...three is the more expensive one, or, one is the more expensive one, I don't really know yet. I guess I could look at the *fare* and maybe figure it out."). Figuring out what *PClass* referred to and how the numbering system for it worked required additional steps to complete.

Authenticity as more work also appeared when Don was asked to select the Movies data set that would provide students with a more authentic experience.

> Don: I would choose [Movies] B, because, it has data that is not necessarily necessary, which I think is-that's how you're going to get data, otherwise they're doing the work for you. Like if you ask your school principal, like hey, can I get some demographic data on the students at our school and the grades that they have? They're not going to just give you two columns, like race, GPA, maybe that's what you wanted to know, but they're not going to do that for you,

As Don described it, there is additional work to do with Movies B to discern what information (i.e., variables) is desirable in the first place. This is the result of having more columns than what is necessary. This implies some searching, selecting, or reducing must be done. Don illustrates this through an imagined scenario of asking the school principal for data. In this imagined scenario, he might want information related to two variables, race and GPA. Assuming the school principal were willing to share the data, "they're not going to just give you two columns". The school is "not going to do that [data reduction] work for you". That there is work that the school will not do implies that there is additional work.

A similar sentiment that authentic means additional work appeared in Roscoe's selection of Movies B as providing a more authentic experience.

> Roscoe: I would choose B, mainly for the volume of data in it, the number of cases, and the fact that it has quite a few attributes, and you have to puzzle down which ones are the ones who want to look at. Also, it looks like there's... I see a couple of NAs in there, I don't see any of the others, I suspect there may be more missing data problems or... outliers or things that have weird attributes on them that you have to deal with, which is always a real problem with real data sets.

There were a few appraised qualities of the data that he noted that suggested more work. One was that there was abundance in the data (which we believe is likely its own aesthetic judgment that we do not detail here. In that aesthetic judgment, large sets are more authentic. Consistent with a KiP model, other knowledge can be cued. In Kiera's example above, 1000 cases in the Titanic B data was suspicious, meaning that a preference for B was superseded by the authenticity as messiness judgment). Abundance appeared in "the number of cases, and the fact that it has quite a few attributes [columns]". Here, the mention of attributes is of current interest. Because there were many attributes, students would have to do the work to "puzzle down".

Another quality of the data was the presence of NAs. These were values that acted comparable to the empty cells that Don noted above in Titanic A data, as this was described by Roscoe as a part of "missing data problems". These were grouped also with "outliers or things that have weird attributes on them that you have to deal with". In having to deal with them, there is added work, whether it is having to make additional decisions (such as whether to exclude those) or caveats (such as the outlier having an influence or inferences being based on less than the total number of cases).

Authenticity as Involving Computing

One other authenticity judgment we discuss came in the form of connection to computing. Data science was something virtually every teacher saw as involving computers and some amount of programming skill. For authentic experiences, some connection to computing seemed connected to a sense of authenticity. For example, Joan immediately noticed the presence of values of "0" and "1" and identified Titanic A as more authentic. She explicitly stated the following:

> Joan: I guess as I teach computer science [in addition to other courses], I like the zeros and ones because it lets me have a conversation about bits and binary [in A].

The main connection Joan made was that it involved an encoding system (binary) that was connected to how computers store information. The presence of "1" and "0" values was mentioned by several teachers for Titanic A in the Survived column. Separate from being an encoding that was associated with computers, the binary values were also seen as being more easily used for computing values. Drake illustrated with the following rationale for choosing Titanic A.

> Drake: I mean I think Titanic A would give them a more authentic experience. If I have to say why, it's that it's quantitative, and I think it's easier to work with quantitative values like, giving the values

survived a value or 1 or 0 is much easier to work with than if it says yes or no, like they could much more easily figure out, okay well how many people survived? Well just add up.

Drake was drawn towards the values being quantitative, although there were quantitative values in all of the data sets. The quantitative values he focused on, however, were “1” and “0”. That encoding for survival would involve a simple matter of adding all values together. There would not be a need to transform written words into numbers that could be manipulated through calculation.

Another way that computing was implicated was in how computers would be necessary to do the work. Roscoe noticed features of the data related to the production of models for real-world phenomena. He elaborated on his noticing of data abundance in his choice of Movies B:

Roscoe: And like I say, Movies A is sort of light on the number of cases. Now, if I was just doing maybe a statistics class, and we weren't actually doing... I could imagine in statistics class, you might not do as much programming. And you know, you're not using it to try to train some model. It may be more appropriate to have fewer things there, that are less confusing and having less data available so that you can actually scan through it and see what the issues are, with, I guess 5,000, there's a chance, you can still stomach going through it, sort of scrolling through everything to see what's there. But definitely 100, people are going to go and look at the entire dataset.

The abundance of data, according to Ramsey, necessitates the use of computational tools and procedures central to data science. This was different from what he considered to be a statistics class where “you might not do as much programming” and the data are not being used “to train some model”. With the larger number of cases in Movies B, which also likely cued the “authenticity as more work” judgment, he suggested it was less likely that students would try to go through thousands of values. He expected that with just 100 data points (Movies A), students would just look at the entire data set to get the information they wanted rather than use computing tools.

Discussion

Thus far, we have examined high school math and statistics teachers’ aesthetic for data sets that support authenticity with data science work. These include sensing authenticity when the data was messy, when using the data required extra work, and when computing had some connection to the work that would be done with the data. These judgments are not exclusive. We suspect from our reviews of the transcripts that they may reinforce one another.

Looking at teachers’ sense of authenticity is consequential for a future where learning scientists will be participating in the work of designing data science educational experiences and for helping educators make judgments on how to provide students with high quality learning opportunities. Our current assumption as learning scientists is that authenticity is desirable. This paper is an initial effort to characterize what that could mean. Having made some observations about how teachers see authenticity in data sets for high school data science instruction, there must also be consideration of the extent to which authenticity is desirable by teachers. It may be that the authentic data sets involve work that the teachers appreciate but struggle to include given the constraints of existing school structures. For instance, data cleaning is a common data science activity and one that takes a substantial amount of data scientists’ time. However, just because it takes much of their time, it may not be something that educators feel is worth proportional instructional time. Indeed, we have begun to examine this in our data. At a later point, we asked teachers to select data sets based on what they thought would be best pedagogically. Of the ten teachers who selected Titanic A as more authentic, five of them said Titanic B would be a better choice pedagogically. However, for the Movies data, the same data set selected for authenticity was also selected by teachers for pedagogical considerations. Examining pedagogical aesthetics and how they complement or conflict with other aesthetics, such as authenticity, remains as work for us to do in the future.

References

Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education* (Vol. 14). Dordrecht, Netherlands: Springer.

Brown, M., & Edelson, D. (2003). *Teaching as design: Can we better understand the ways in which teachers use materials so we can better design materials to support their change in practice?* Design Brief. Northwestern University. Evanston, IL.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age.* Hillsdale, N. J.: Lawrence Erlbaum.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2&3), 105-225.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843-900.

diSessa, A. A., Levin, M., & Brown, N. J. S. (Eds.). (2016). *Knowledge and Interaction: A Synthetic Agenda for the Learning Sciences*. New York, NY: Taylor & Francis.

diSessa, A. A., & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10), 1155-1191.

diSessa, A. A., Sherin, B., & Levin, M. (2016). Knowledge Analysis: An Introduction. In A. A. diSessa, M. Levin, & N. J. S. Brown (Eds.), *Knowledge and Interaction: A Synthetic Agenda for the Learning Sciences* (pp. 30-61). New York, NY: Taylor & Francis.

Domahs, F., Moeller, K., Huber, S., Willmes, K., & Nuerk, H.-C. (2010). Embodied numerosity: Implicit hand-based representations influence symbolic number processing across cultures. *Cognition, 116*(2), 251-266. doi:https://doi.org/10.1016/j.cognition.2010.05.007

Edelson, D. C. (2004). My World: A case study in adapting scientists' tools for learners. In Y. Kafai, W. Sandoval, N. Enyedy, A. Nixon, & F. Herrera (Eds.), *Proceedings of the 6th international conference on Learning sciences*. Santa Monica, CA: Lawrence Erlbaum Associates.

Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 335-354). New York: Cambridge University Press.

Gould, R., Machado, S., Johnson, T. A., & Molynoux, J. (2018). *Introduction to Data Science v 5.0*. Los Angeles: UCLA Center X.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of Learning from a Modern Multidisciplinary Perspective* (pp. 89-120). Greenwich, CT: Information Age Publishing.

Lave, J., Murtaugh, M., & de la Rocha, O. (1984). The dialectic of arithmetic in grocery shopping. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: It's development in social context* (pp. 67-94). Cambridge, MA: Harvard University Press.

Lee, V. R., & Delaney, V. (2021). *What Is Being Covered in Standalone Secondary School Data Science Curricula?* Paper presented at the 2021 Annual Meeting of the American Educational Research Association. [Virtual]

Lee, V. R., & Wilkerson, M. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects*. Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C.

Murata, A. (2004). Paths to Learning Ten-Structured Understandings of Teen Sums: Addition Solution Methods of Japanese Grade 1 Students. *Cognition and Instruction, 22*(2), 185-218. doi:10.1207/s1532690xci2202_2

Newman, G. E. (2019). The Psychology of Authenticity. *Review of General Psychology, 23*(1), 8-18. doi:10.1037/gpr0000158

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109*(2), 330-357. doi:10.1037/0033-295X.109.2.330

Philip, T. M. (2011). An "Ideology in Pieces" Approach to Studying Change in Teachers' Sensemaking About Race, Racism, and Racial Justice. *Cognition and Instruction, 29*(3), 297-329. doi:10.1080/07370008.2011.583369

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., . . . Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632-654. doi:https://doi.org/10.1002/tea.20311

Sherin, B., Krakowski, M., & Lee, V. R. (2012). Some assembly required: How scientific explanations are constructed in clinical interviews. *Journal of Research in Science Teaching, 49*(2), 166-198. doi:10.1002/tea.20455

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the learning sciences, 3*(2), 115-163.

Wagner, J. F. (2006). Transfer in pieces. *Cognition & Instruction, 24*(1), 1-71.

Wilkerson, M. H., & Polman, J. L. (2020). Situating Data Science: Exploring How Relationships to Data Shape Learning. *Journal of the learning sciences, 20*(1), 1-10. doi:10.1080/10508406.2019.1705664