The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2398-5348.htm

Received 20 June 2023
Revised 17 September 2023
Accepted 2 October 2023

Teen-adult interactions during the co-design of data literacy activities for the public library: insights from a natural language processing analysis of linguistic patterns

Leanne Bowler and Irene Lopatovska
School of Information, Pratt Institute, New York, New York, USA, and
Mark S. Rosin
Department of Math and Science, Pratt Institute, New York, New York, USA

Abstract

Purpose – The purpose of this study is to explore teen-adult dialogic interactions during the co-design of data literacy activities in order to determine the nature of teen thinking, their emotions, level of engagement, and the power of relationships between teens and adults in the context of data literacy. This study conceives of co-design as a learning space for data literacy. It investigates the teen–adult dialogic interactions and what these interactions say about the nature of teen thinking, their emotions, level of engagement and the power relationships between teens and adults.

Design/methodology/approach – The study conceives of co-design as a learning space for teens. Linguistic Inquiry and Word Count (LIWC-22), a natural language processing (NLP) software tool, was used to examine the linguistic measures of Analytic Thinking, Clout, Authenticity, and Emotional Tone using transcriptions of recorded Data Labs with teens and adults. Linguistic Inquiry and Word Count (LIWC-22), a natural language processing (NLP) software tool, was used to examine the linguistic measures of Analytic Thinking, Clout, Authenticity and Emotional Tone using transcriptions of recorded Data Labs with teens and adults.

Findings – LIWC-22 scores on the linguistic measures Analytic Thinking, Clout, Authenticity and Emotional Tone indicate that teens had a high level of friendly engagement, a relatively low sense of power compared with the adult co-designers, medium levels of spontaneity and honesty and the prevalence of positive emotions during the co-design sessions.

Practical implications – This study provides a concrete example of how to apply NLP in the context of data literacy in the public library, mapping the LIWC-22 findings to STEM-focused informal learning. It adds to the understanding of assessment/measurement tools and methods for designing data literacy education, stimulating further research and discussion on the ways to empower youth to engage more actively in informal learning about data.

Originality/value – This study applies a novel approach for exploring teen engagement within a co-design project tasked with the creation of youth-oriented data literacy activities.

Keywords Data literacy, Youth, Co-design, Participatory design, Natural language processing, Informal learning, LIWC-22

Paper type Research paper

Information and Learning Sciences
Vol. 125 No. 3/4, 2024
pp. 252-269

2398-5348
DOI 10.1108/ILS-06-2023-0076

The authors thank the teens and librarians who participated in this research. Their contributions are invaluable. Many thanks as well to Jiacheng Chen for his contribution to the project. This research was made possible with support from the National Science Foundation (Award #2005608).

Introduction

We live in a world of data-gathering, data-surveillance and data-driven decision-making by both humans *and* machines, necessitating an urgent need for a data-literate citizenry. Today's young people, more than any previous generation, have a personal stake in their ability to function with data, being themselves the subjects *of* data science and the most thoroughly measured, tracked and analyzed generation (Taylor and Rooney, 2016). The *Data Literacy with, for, and by Youth* research project addressed this critical need through an informal learning design study, where we worked alongside teens to design, develop and test potential after-school data literacy activities that they, the teen participants, thought would be suitable for a teen audience at the public library. This article investigates the teen–adult dialogic interactions that occurred during the design sessions (called Data Labs in this study) and what these interactions say about the nature of teen thinking, emotions, level of engagement and power relationships between teens and adults. We ask, *What do teens' dialogic interactions with each other and with adults (spoken or in group chats in Zoom) reveal about their thinking, emotion, engagement and their roles in the co-design of data literacy activities?*

Over the course of two years and four series of Data Labs which included 24 after-school data labs (online and in-person), the *Data Literacy with, for, and by Youth* research project explored three themes:

(1) teens' critical thinking about data;

(2) their engagement with data in an after-school setting; and

(3) the role of teens in the co-design of data literacy activities at the library.

To address these themes, the project iteratively designed and developed, alongside 48 teen participants, a collection of adaptable informal learning activities for public libraries centered around data literacy. The teens tried out, made improvements to or designed their own data literacy activities for the public library, such as, for example:

- online quizzes about digital data using Kahoot, a game-based learning platform;
- a prototype for a board game about digital data and privacy modeled on Monopoly (which teens called *Data-opology)*;
- a storyboard for a DC comic about data privacy and surveillance;
- a "build" project for Minecraft using Motor Vehicle Collision Data; and
- a community-improvement project for a park, using New York City Park Crime Statistics.

Throughout the design process, there was much talk about data – what it is, how it is embedded in teen lives, what data rights teens have, how teens might use their knowledge of data in the community and what types of activities would work best with other teens. In this article, we use the transcripts of these conversations to explore the project's driving themes of critical thinking, engagement and the role of teens in the co-design of data literacy activities, the purpose of which is to add to our understanding of assessment/measurement tools and methods for learning about data literacy. Through the language of teens, we explore the teens' thoughts, feelings and the ways they connected with each other and with adults are revealed. We apply the Linguistic Inquiry and Word Count tool (also known as LIWC-22, pronounced "luke"), a widely used computer-based text and speech analysis tool for the psychological study of everyday language, and we analyze recorded interactions through the variables of *Clout* (social status, confidence or leadership), *Analytic Thinking* (formal, logical and hierarchical thinking patterns), *Authenticity* (spontaneity and honesty) *and Emotional Tone* (positive and negative emotion).

This article does not set out lesson plans for data literacy at the library. Indeed, our purpose here is not to describe the Data Labs in terms of teaching techniques, activities or procedures (readers interested in step-by-step guidelines can refer to Bowler *et al.*, 2022a, 2022b). Nor was our purpose to evaluate learning outcomes. Rather, in the context of data literacy education, we are interested in the *process of learning and its manifestations through language*. We have, therefore, analyzed the teen–adult conversations *during the co-design sessions*, looking for evidence of engagement, areas of strength and possible red flags for practitioners and researchers to consider when co-creating, alongside teens, data literacy activities in informal learning settings such as the public library.

Background

Data literacy at the public library

Data literacy is a complex array of skills, knowledge and humanistic reasoning to be applied throughout the data life cycle. It includes the ability to critique data practices; to contextualize data to broader contexts such as platforms, cyberinfrastructure and society; and to find meaning in data beyond statistical and mathematical arguments (Finzer, 2013, p. 5; Deahl, 2014; Tygel and Kirsch, 2015; Gray *et al.*, 2018; Acker and Bowler, 2018; Bowler *et al.*, 2017). Data literacy has cognitive, affective and behavioral aspects (Chi *et al.*, 2018). Interest and engagement are key to all learning (Hidi, 1990; Silvia, 2008), including learning experiences with data (Bowler *et al.*, 2020). Young people, through the application of participatory design, can show us what is interesting and meaningful to them when it comes to learning about data at the public library (Bowler *et al.*, 2021; DiSalvo *et al.*, 2012), a key factor in their engagement with data.

Data literacy education for teens in public libraries is an emerging area of library programming, often under the guise of digital literacy, privacy and internet safety. Public libraries, whose mission is to provide access to information, also inform the public about civic and government data, such as Census Data or local repositories like NYC Open Data and the Western Pennsylvania Regional Data Center, through library activities such as open data jams, hackathons and workshops. Research in after-school learning shows us that there are many ways public libraries can approach something as complex as data literacy in the informal learning setting. A meta-analysis of 73 after-school programs found that the personal and social development of youth were enhanced through their participation in structured, voluntary, out-of-school activities (Durlak and Weissberg, 2007). Research in after-school programs has also shown that they are a rich environment for learning, providing the motivational environment and social arrangements that help propel interest and engagement in a topic, hobby or new skill (Larson, 2000; Csikszentmihalyi, 1990; Csikszentmihalyi and Kleiber, 1991; Csikszentmihalyi and Larson, 1984). Larson, an early advocate for after-school, community-based programs for youth, argues that adults interested in positive youth development should "give youth activities equivalent status to school, family, and peers as a focal context of development" (Larson, 2000, p. 178).

This study into data literacy was grounded in a framework for informal STEM learning, which is typically driven by learner interests and motivation and is contextually relevant and personal, voluntary and open-ended (Falk, 2001; Falk and Dierking, 2000). The value of informal science learning in out-of-school environments has been acknowledged by the National Research Council, in its report, National Research Council (2009). The report sets out goals for science programming in out-of-school environments (such as, e.g. data literacy activities at the library), highlighting six strands (or outcomes) of science learning. The six strands represent the ideal set of outcomes for informal science learning but Strand 1 – learners who engage with science in informal environments should experience excitement, interest and motivation to learn about phenomena in the natural and physical world – is particularly relevant to this study because, given the "drop-in" nature of informal learning at the library, the need to quickly absorb people into the learning process is central to success. Strand 1 therefore speaks to the need for a deeper understanding about how to generate excitement and motivation and build on the learner's personal interests. In short, how to engage learners in STEM. In this study, we addressed Strand 1 by using data sets and data problems that would "make data personal and close to support emotional engagement" (Bilstrup *et al.*, 2022).

In this article, we explore manifestations of engagement in informal learning about data literacy based on the linguistic analysis of conversations that took place during co-design sessions. Using co-design techniques to build a youth-oriented model of data literacy activities for the library offers both opportunities and challenges. Methods of cooperative inquiry, an approach commonly practiced in technology design but also associated with action research, can be applied to the co-design of youth-focused library services, with teens and librarians working collaboratively (Druin, 1999; Druin, 2002; Yip *et al.*, 2017; Subramaniam, 2016). A model of participatory librarianship, informed by the four dimensions of interaction described in their earlier work – *facilitation*, *relationship-building*, *design-by-doing* and *elaborating together* – was set out by Yip *et al.* (2019). In this model, the librarian role reflects degrees of interaction, from supportive to co-design. Educators in Library and Information Science (LIS) have argued that the techniques of participatory design are an essential skill set for youth librarians (Subramaniam, 2016) and indeed, many public libraries have Teen Advisory Boards facilitated by a youth librarian, where youth design programs and policies for the library.

Navigating the delicate balance between the need to convey complex STEM concepts associated with data versus the expectations that teens bring to the public library necessitates a level of "conscious co-design" (Bowler *et al.*, 2021). We take this responsibility further, by delving deeper into the nature of the rapport between teens and adults during the co-design of data literacy activities, using linguistic analysis tools and the transcripts from co-design sessions. Although this method takes a deep dive into what people said, we note that the teens' confidentiality has been protected and their identities made anonymous throughout this study through the removal of identifying details and the use of pseudonyms.

Natural language processing

Language is a medium through which human beings express their thoughts, feelings and needs. Languages are highly patterned so that they can be understood widely, allowing for systemic analysis of spoken and written words. Using the psychometrics of words, this project opens a window on teen experiences in the *Data Literacy with, for, and by Youth* research project, which ran a series of co-design activities, where teens created data literacy experiences with, for, and by other teens. To analyze teens' thinking, emotion and engagement, this project applied the LIWC-22, a widely used computer-based text analysis tool for the psychological study of everyday language. LIWC-22 is one of a family of linguistic analysis tools that apply natural language processing (NLP) techniques.

NLP systems combine elements of computer, artificial intelligence and linguistics to develop systems that can process and analyze data consisting of human language (Sowmya, 2022). Although NLP still faces many challenges related to the ambiguity and creativity of human language, it has also emerged as a useful tool for analyzing large data sets consisting of text or recorded speech to generate word clouds, analyze sentiment and categorize language into topics or themes. It has the advantage of being able to quickly and efficiently (once the data is cleaned) extract broad statistically significant trends from data sets that might, otherwise, be too labor intensive to analyze. This has had a wide range of research applications in recent years, from using NLP models to analyze survey data relating to job satisfaction (Speer *et al.*, 2022) to the detection of suicidal ideation (Arowosegbe and Oyelade, 2023) and early signs of Alzheimer's (Diaz-Asper *et al.*, 2022).

One subfield of NLP, sentiment analysis, has emerged in recent years as a tool for identifying and studying subjective data points such as emotions underlying written or spoken responses. With the LIWC-22 tool, for example, *positive and negative tone* is a form of sentiment analysis. Sentiment analysis is used in a range of fields, including marketing to evaluate the nature of engagement with social media posts (Ma *et al.*, 2022), and in therapeutic contexts to quantify emotions underlying patient responses (Tanana *et al.*, 2021; Leung *et al.*, 2021). In the information science field, sentiment analysis has been proposed as an effective method for identifying and analyzing user engagement and satisfaction with the library.

As public facing institutions, libraries regularly collect user feedback in the form of user satisfaction surveys with the stated purpose of working to better meet public needs. Recent studies have suggested that the use of sentiment analysis to analyze such survey responses has the potential to "unveil latent emotions" underlying user feedback and allow for a more wide-ranging analysis of responses outside of the "confined scale" of the survey parameters (Papachristopoulos *et al.*, 2019). NLP sentiment analysis has also been proposed as a potentially valuable tool for analyzing library reference chat transcripts for sentiment patterns such as satisfaction and confusion to adjust library services to better meet user needs (Ozeran and Martin, 2019).

Other recent examples of the use of NLP tools in the field of information science include the analysis of text data from reference chat transcripts with the goal of categorizing the types of questions asked by library patrons to predict future questions, more efficiently channeling questions to appropriate library staff, and informing library staff about a need for further support and guidance (Wang, 2022; Ozeran and Martin, 2019).

Linguistic inquiry and word count-22

There are many approaches to text analysis, from the inductive qualitative approach of open coding as used in Grounded Theory, to computerized natural language analysis based on statistical algorithms, as discussed above. In this study, we applied LIWC-22, a computer program that looks for and counts words in psychology-related categories and characterizes text by emotion, thinking style and the nature of social interaction. According to the creators of this linguistic analysis tool, over 20,000 scientific articles have been published using LIWC. The LIWC-22 tool reviews all the words in a text, coding them according to categories in the LIWC dictionaries, and then outputs the percentage of words in a text that fall into linguistic, psychological and topical categories. In the area of teens and STEM learning, the LIWC tool has been used to evaluate engagement in project-based learning with students in grade seven and eight (Lee *et al.*, 2018), assess outcomes in after-school Makerspaces through linguistic analysis (Oliver *et al.*, 2021), and to examine student reflections in a STEM enrichment summer program (Matheson *et al.*, 2017).

At the heart of the LIWC-22 tool is the dictionary and a text processing engine. Human judges have developed and validated 100 dimensions of text within the dictionary that are associated with basic emotional and cognitive dimensions, looking for "goodness of fit." The initial selection of word candidates has been drawn from dictionaries, thesauri and word lists from experts. Groups of three to four judges then independently rated whether the word candidates fit into broader word categories. The dictionaries have been revised and validated multiple times over the course of three decades, starting in 1992 and most recently in 2022. A "Test Kitchen" corpus of 31 million words, including a range of textual formats, such as blog posts, social media, conversation, news sources, novels and student writing, then serves to validate the measurement instruments as well as establish benchmarks for various textual formats. In the most recent version of LIWC (2022), language reflective of the discourse common to social media and SMS styles of communication (i.e. short messaging services) has been incorporated. For more details about the formulation of the dictionaries, see Boyd *et al.* (2022a, 2022b) and Pennebaker *et al.* (2007).

There are 100 categories and subcategories, arranged hierarchically, within the LIWC-22 dictionary, including four broad summary language measures of *Analytic Thinking*, *Clout, Authenticity* and *Emotional Tone*. For the purposes of this study, we chose to use these four broad summary language measures to reveal engagement and the interpersonal roles of teens and adults in the co-design process of data literacy activities. While more specific categories and subcategories in the LIWC-22 dictionary would allow for a targeted approach (e.g. looking for words that suggest tentativeness, anger, past, present or future focus, to name a few), they would not give as complete a picture as the four summary language categories, because the broader measures contain all words and concepts in the dictionary. Analysis via the four broadest categories, and therefore, the most inclusive approach linguistically speaking, provides the most statistically reliable results – an important consideration when not dealing with Big Data. Furthermore, the summary language measures interpret the meaning of words, going beyond a basic, raw count of words and applying psycholinguistic theory. The summary language categories of *Analytic Thinking*, *Clout*, *Authenticity* and *Emotional Tone* are defined below. For the complete list of categories, see Boyd *et al.* (2022a, 2022b, pp. 11-12).

Methodology

In this section, we elaborate on the data sources and methods of data cleaning, processing and analysis, including definitions of variables.

The data set consists of transcripts from audio recordings from online and in-person sessions and written chat conversations in Zoom that took place during 24 Data Labs with 48 teen participants, in four separate design teams, at the Brooklyn Public Library from Spring 2021 to Spring 2022. Teen participants worked alongside researchers and librarians in a co-design process developing data literacy activities for the library. Each design session ran for 90 min and was audio recorded. Nineteen Data Labs were online, conducted via the Zoom video conferencing platform, thus allowing for the capture of chat conversations as well. Five were in-person at the library. Transcripts from the Data Labs were aggregated into two sets of text files compatible with LIWC-22 – one set for audio recordings, the other for chat – to provide higher accuracy in semantic analysis. The first step in data cleaning was to determine speaker type through a close reading of the transcripts. The speakers were categorized as teens, session leaders and librarians. All identifying details relating to individual participants – both the teens and adults – were removed from the transcripts.

Algorithmically, the data set was then aggregated based on three criteria – speaker type, the data lab session number and the design team ID – while data such as timestamps and URLs were removed to avoid confusion during semantic analysis. The data set was then processed through the LIWC-22 linguistic analysis tool (see Figure 1 below, showing a LIWC-22 analysis). We then compared the results with benchmarks established in the LIWC-22 "Test Kitchen," focusing on the summary language categories of *Clout*, *Analytic Thinking*, *Authentic* and *Emotional Tone* to determine the significance, or lack thereof, of these sentiments in the data labs. Each of the measures for the summary language categories are algorithms derived from various LIWC variables based on previous empirical research in linguistics such as, for example, the role of pronouns in expressing positionality in a relationship (Boyd and Schwartz, 2021; Boyd and Pennebaker, 2016).

We chose the four summary language categories because of their potential to reveal the teens' cognitive and emotional processes, aspects of their engagement with data and their role in the co-design of data literacy activities. The summary language categories in LIWC-22 were created through algorithms derived from variables in the LIWC Dictionary and based on previous empirical research in psycholinguistics. Rather than a basic word count, the measures for summary categories reflect the psychometrics of words and are standardized scores that have been converted to percentiles ranging from 1 to 99. *Analytic Thinking* is the metric of logical, formal thinking, as reflected in linguistic expressions in conversation and texts. Analytic Thinking is in the style of school discourse and tends to be rewarded in academic settings. Ranked on a scale of 0–99, language scoring low in Analytic Thinking tends to be viewed as more friendly and casual. *Clout* is an expression of status and leadership. "People high in clout speak with a sense of certainty and authority, whereas people low in confidence sound more tentative and uncertain" (Jordan *et al.*, 2019). Ranked on a scale of 0–99, a high Clout number suggests expertise and confidence while a low Clout number suggests a more humble, cautious, even anxious attitude. The category labeled *Authentic* reflects a perceived honesty and genuineness, but also, less self-filtering. Low authenticity could indicate someone is socially cautious. *Emotional Tone* relates to the degree of positive or negative tone. A high score on emotional tone is associated with a more positive, upbeat style; a low score reveals greater anxiety, sadness or hostility.

In the final stage of analysis, the LIWC-22 results were further explored in terms of their standard deviation and *Z*-score, a statistical measure that reveals the deviation of our test results from the benchmark mean, calculated based on the "Test Kitchen" benchmark mean and standard deviation. In this study, we compared our results with the "conversational" LIWC benchmark: a corpus of 1,000 texts, each originally between approximately 100 and 10,000 words (Boyd *et al.*, 2022a, 2022b). This examination of *Z*-scores allowed for the identification of statistically significant differences in comparison to the "Test Kitchen" benchmark.

Findings

Below we discuss the findings with regard to the four summary language categories – *Analytic Thinking*, *Clout*, *Authentic* and *Emotional Tone*. The LIWC output can be viewed through the lens of linguistics, statistics and psychology (Interpreting the LIWC Output, 2022).

Summary language categories: a detailed description

Analytic thinking. The summary language category of Analytic Thinking "captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns."


wi22_chat_teen

Filename	Segment	WC	Analytic	Clout	Authentic	Tone
wi22_chat_teen.txt	1	2122	30.77	49.31	42.04	92.08

Figure 1.
Screenshot of an LIWC-22 linguistic analysis of the teen chat transcripts from Winter 2022

Note: Showing measures for summary language categories
Source: Figure by authors

People tend to use analytic language in more formal settings, such as classroom environments, whereas in informal learning they would use language that is more intuitive and personal. It is worth noting that language that scores high in *Analytic Thinking* is the type of speech that is rewarded in academic settings. In contrast, language that scores low in *Analytic Thinking* is interpreted as more friendly, intuitive and less cold – exactly the kind of language one might hope to hear in a voluntary, after-school activity with teens at the public library.

In our LIWC analysis, the teen's mean scores for *Analytic Thinking* across all Data Labs are low (19.17 for audio; 34.17 for chat), suggesting that teens communicated in a friendly, casual manner (see Table 1). This type of interaction would be prized in the informal setting of a library program but also in the context of co-design, where open discussion and consensus are valued.

Interestingly, both the teen audio and chat discourse in the data literacy sessions were more analytically focused than the Test Kitchen's own benchmark for "Conversation," perhaps because conversation during our data labs did involve new insights and learning, albeit in a casual, after-school setting (the *Z*-scores for audio and chat are 1.39 and 2.46, respectively). However, it is important to note that the Test Kitchen's "Conversation" benchmark is specifically characterized by its casual, low-analytic nature, and in fact it is particularly low in *Analytic Thinking* scores compared with other Test Kitchen benchmarks. Although there are other types of benchmarks from the Test Kitchen, we chose to use the "Conversation" benchmark for consistency and to reflect the nature of the discourse in our transcripts, which was live, interactive group conversation, written or verbal. Comparing our LIWC scores with other benchmarks would reveal a significantly lower level of *Analytic Thinking*, thus supporting our findings. For example, comparing our data with text gleaned from technical college application essays in the analytic category, these data have a significantly lower *Z*-score: −6.15 and −5.26 in the Zoom audio and Zoom chat categories, respectively.

As one teen noted in an exit survey, "The most important thing to keep in mind is to keep it fun and not too structured," seemingly calling for a discourse environment that is distinctly "not school." The low *Analytic Thinking* LIWC score for teens suggests that this was indeed the case and that a high level of friendly engagement was experienced. This is important when considering the nature of informal STEM learning, which has a different structure than that of formal, classroom learning. Informal *STEM* learning is typically driven by learner interests and motivation that is contextually relevant and personal, voluntary and open-ended (Falk, 2001; Falk and Dierking, 2000; National Research Council, 2009). A low *Analytic Thinking* score might be distinctly advantageous in such an environment.

Clout. *Clout* is a measure of "relative social status, confidence, or leadership that people display through their writing or talking" (Jordon *et al.*, 2019). In their study of linguistic

	Teens			
	Audio-teens		Chat-teens	
LIWC measure and data source	LIWC score	*Z*-score	LIWC score	*Z*-score
Analytic thinking – all data labs	20.87	1.39	29.66	2.46

Note: LIWC score: Numbers are standardized scores that have been converted by LIWC to percentiles ranging from 1 to 99
Source: Table by authors

Table 1.
Analytic thinking – teens – all series

patterns of politicians, a group of people commonly accepted as high in power, Jordon *et al.* (2019) found that "people high in clout speak with a sense of certainty and authority, whereas people low in confidence sound more tentative and uncertain."

The study of *Clout* is relevant for an intergenerational, co-design project with youth such as this because it illustrates the balance of power between members of the design team. Because *Clout* can reveal power relationships, in this analysis we include data for *both* teens and adults. In the data, we see that the teens' *Clout* scores are lower than that of the adults, for both the audio and chat transcripts, suggesting a clear power dynamic between teens and adults (see Table 2). LIWC scores for summary language categories such as Clout are ranked on a scale of 1–99. Notably, the adult's LIWC scores for the audio transcripts are 91.20, illustrating that, even though the language used was casual, friendly and generally positive, adults held court during the verbal conversation. This, even though the language used was casual, friendly and generally positive. In chat, adults scored somewhat lower (77.45), but this is still a much higher number than the teen's 43.96. Notably, the "clout gap" was greatest during the in-person sessions held in the library.

Authentic. Authenticity in linguistic expression speaks to spontaneity, honesty and a reduction of the filters we often apply to our speech. A high score in *Authentic* reflects a spontaneous conversation between friends. This is relevant for a participatory design project that aims to hear the honest views of young people in the design process.

The LIWC scores for teens reveal a modest level of *Authentic* in their linguistic expression, the average across all series and all modalities (online and in-person at the library) being 61.31 (see Table 3). This score lies in the center of the 1–99 scale – not high, not low. Overall, the LIWC scores do fall in line with those for the Test Kitchen's

Table 2.
Clout – comparison of teen and adult transcripts – all series

LIWC measure and data source	Teens				Adults			
	Audio		Chat		Audio		Chat	
	LIWC score	*Z*-score	LIWC score	*Z*-score	LIWC score	*Z*-score	LIWC score	*Z*-score
Clout – all data labs	50.11	0.50	43.96	0.24	91.20	2.27	77.45	1.68

Note: LIWC scores: Numbers are standardized scores that have been converted by LIWC to percentiles ranging from 1 to 99
Source: Table by authors

Table 3.
Authentic – teen transcripts by series

Data source for LIWC measure of authentic	Modality	Teens			
		Audio-teens		Chat-teens	
		LIWC score	*Z*-score	LIWC score	*Z*-score
Data Labs: Spring 2021 cohort	Online	56.39	−0.66	58.34	−0.56
Data Labs: Fall 2021 cohort	Online	58.68	−0.54	70.50	0.05
Data Labs: Winter 2022 cohort	Online	62.72	−0.34	44.33	−1.26
Data Labs: Spring 2022 cohort	In-person at the library	81.77	0.61	No chat	
Average AUTHENTIC across all series		64.89	−0.23	57.72	−0.59

Note: LIWC scores: The numbers are standardized scores that have been converted by LIWC to percentiles ranging from 1 to 99
Source: Table by authors

"Conversation" category, suggesting that the authenticity of the teen's linguistic expressions is on par with typical conversation.

Emotional tone. The summary language category of *Emotional Tone* summarizes positive and negative emotion words in written text and compiles them into a single summary variable – numbers below 50 suggest a more negative emotional tone (LIWC Analysis, 2022; Cohn *et al.*, 2004).

While some LIWC analyses can focus on specific emotion variables, such as anger or sadness, for example, it does so as a simple raw count of words and percentage of all words in the text. To gain a broader overview of emotion in the discourse, we used the summary language category of *Emotional Tone* because it interprets the meaning of words *and* the raw number count into a more revealing score on a scale of 1–99. The higher number, the more positive the tone.

With this in mind, we see that the teen's *Emotional Tone* overall (averaged across modalities and series) is modestly high (74.35), evidence of a general positive feeling throughout all series of the Data Labs. Notably, the *Emotional Tone* communicated through chat transcripts is significantly higher than that of the audio transcripts – the average score of the chat transcripts being 91.47 out of 99 (note that the in-person series did not use the chat feature in Zoom) (Tables 4 and 5).

Discussion

This study asked the question, *What do teens' dialogic interactions with each other and with adults (spoken or in group chats in Zoom) reveal about their thinking, emotion, engagement with data and their roles in the co-design of data literacy activities?* The question was explored using the LIWC-22 linguistic analysis tool and through the lens of the summary language dimensions of *Analytic Thinking*, *Clout*, *Authentic* and *Emotional Tone*. The study shows that there was a power dynamic at play between the teen participants and the adults which

	Teens			
	Audio-teens		Chat-teens	
LIWC measure for data source	LIWC score	*Z*-score	LIWC score	*Z*-score
Average emotional tone across all series of data labs	57.23	−0.05	91.47	1.19

Notes: LIWC score: The numbers are standardized scores that have been converted by LIWC to percentiles ranging from 1 to 99. Numbers below 50 suggest a more negative emotional tone
Source: Table by authors

Table 4. Emotional Tone – teen transcripts – all series

Data source for LIWC measure of emotional tone	Modality	Teens' LIWC score Audio-teens	Chat-teens
Data Labs: Spring 2021 cohort	Online	75.91	94.16
Data Labs: Fall 2021 cohort	Online	45.26	87.34
Data Labs: Winter 2022 cohort	Online	56.02	92.92
Data Labs: Spring 2022 cohort	In-person at the library	51.73	No chat

Notes: LIWC scores: The numbers are standardized scores that have been converted by LIWC to percentiles ranging from 1 to 99. Numbers below 50 suggest a more negative emotional tone
Source: Table by authors

Table 5. Emotional Tone – teen transcripts by series

one might naturally expect to impact engagement. Counterintuitively it did not. There is strong evidence of engagement, with the teens establishing a friendly rapport, with each other and the adults, and expressing positive feelings about the design process. In participatory design, where participants' honest opinions are necessary to build a product that reflects the user's needs, authenticity is critical. In this study, the teens were modestly open in their linguistic expressions but did not quite reach the threshold of "trusted friend." An interesting anomaly, however, was that the in-person Data Labs registered much higher on the scale of *Authentic*, suggesting that the co-design of data literacy activities is best achieved in person rather than online. These results lead us to reflect on the broader implications to arise from this inquiry:

- *Clout and intergenerational co-design*: There is abundant research in participatory research with children and youth, providing models in support of intergenerational design teams (e.g. the work of Hart, 1992; Druin, 1999, 2002; Large *et al.*, 2006; Yip *et al.*, 2017, 2019). Even so, the *Clout* scores in this study call into question whether a true equity between adult and youth design partners can exist, and points to age-old disparities between adults and youth. We call this the "clout gap." In the case of the authors, we were well-aware of our own power relationship with the teens throughout the study, even to the point of completing a self-reflection after many of the Data Labs and writing about the need for "conscious co-design" and the "self-reflective and deliberative planning for participation in co-design, particularly on the part of adults working with vulnerable populations" (Bowler *et al.*, 2021). In the event, most likely the *Data Literacy with, for, and by Youth* project only reached the mid-range of participation on Hart's (1992) *Ladder of Participation* – identified by Hart as "Adult Initiated, Shared Decisions."

The real question is whether an intergenerational design team on a research project that was conceived of and initiated by adults can ever truly achieve an equal level of power between adults and youth. With this question in mind, it is interesting to compare the low *Analytic Thinking* scores of teens (demonstrating a casual, friendly atmosphere) with the seemingly large gap between Teen *Clout* and Adult *Clout*. The teen's awareness that they were participating in an "adult project" – a project that they clearly did not initiate – could have led to teen language that deliberately expressed less *Clout*. Nevertheless, the possible negative effects of *Clout* on teen engagement seem to have been mitigated by other variables at play in the Data Lab sessions, which, together, helped facilitate less formal, more friendly language. These variables might have included the adults' high interest in teen perspectives, the inclusion of participatory practices throughout the sessions, the low stakes learning environment (no grading!), coupled with the after-school environment of the library (not school!):

- *Environments for authentic interaction with teens*: As we noted above, there was a modest level of *authenticity* in the teen's linguistic expression and we suggest there are two factors that account for this: first, the teens on each design team did not know each other before their participation in the project. Although *friendly* during the sessions, they were not friends. Nor had they met the adults previously. Furthermore, the entire project took place during the COVID-19 pandemic, during which time many teens' social connectivity and skills were adversely affected. Given this situation, it would seem normal that a true, friendship-driven conversation would be less likely. The LIWC-22 scores for the audio transcripts recorded during the in-person sessions at the library are quite high (81.77) in comparison to the lower scores for audio transcripts recorded during the online sessions. The scores from the in-person sessions thus raise the overall average of all scores. What was it about the in-person sessions at the library that made linguistic expression more authentic than those of the online conversations? Is the online environment (during a global pandemic, no less) not the right environment for deep and honest discourse? Did small delays in online exchanges and the participants' ability to "hide" during Zoom sessions subsequently make teen responses less spontaneous?

- *Positive experiences for teens*: As we report, the teens exhibited a positive *Emotional Tone* throughout all series of the Data Labs. Interestingly, the positive tone of chat transcripts was higher than those of the audio transcripts (both in-person and online). This data point supports the observational data reported earlier in this same study (Bowler *et al.*, 2022a, 2022b), where the researchers noted teens' higher engagement with online group chat versus verbal interaction. Chat was the teen's preferred form of discourse, rather than verbal conversation, with many of the teens turning off their cameras during Zoom sessions and silent during group conversations. We wondered why chat produced a more positive *Emotional Tone* than verbal conversation (although both were positive). Perhaps turning off cameras and stepping outside of the conversation was a distancing technique influenced by the teen's fatigue of online learning during the COVID-19 pandemic. This is an understandable response but at the same time, it may have affected the level of positive emotional tone in the verbal conversations. An alternative explanation might be that, seeing chat as their own teen space, they felt more positive there (in chat) than when engaged in a verbal conversation ostensibly led by an adult. It is interesting to compare the teen's low sense of *Clout*, discussed above, with their positive *Emotional Tone*. Given that the data literacy project was an adult-initiated endeavor, low confidence in terms of their power positioning may not have impacted the teen's general pleasure, fun and engagement. Like school, work and in their families, teens are usually not in positions of power – that is the norm for them – but that does not mean they do not experience something good and positive. Having said that, the authors advocate for both *Clout* and a positive *Emotional Tone* and call for more discussion about how to empower youth in participatory design.
- *Methodological implications*: The method presented here offers a novel approach to assessing informal learning activities and would be of value to practitioners and researchers interested in measuring learner's engagement. Engagement is more than simple participation. It is "the extent to which young people are focused on and excited about the activities they are participating in at a particular point in time and space" (Walker *et al.*, 2005, p. 403). Arguably, engagement is key in the informal learning environment of the library, given the voluntary nature of after-school activities. In this study, we measured teen engagement via the linguistic properties of conversation during design sessions for data literacy activities and demonstrated how a NLP tool, in this case LIWC-22, can reveal teen engagement. There are, of course, many ways to assess engagement. For example, Bowler *et al* (2020) applied inductive methods by open coding observation notes from 27 data literacy workshops, whereas Lee (2021) used electrodermal activity data and videos to obtain psychophysiological data about youth engagement in maker activities at the library, and Ryu and Lombardi (2015) combined critical discourse analysis with a social network analysis. In this article, we point the way to a new approach, applying NLP techniques – specifically the LIWC-22 linguistic tool – to explore a series of after-school data literacy activities with young people. Our goal was to reveal aspects of the teen's role in co-design, and critically, their engagement with

data literacy activities, and by extension, their engagement with data. Applying NLP was a good *first step* for this project, given the body of research data that we had to process. We found the NLP techniques to be valuable because they pointed the way toward broad themes that reveal areas of opportunity or caution, deserving of a deeper thinking about how to plan for teen–adult interaction in the co-design of STEM activities such as data literacy. Although NLP tools have shown to be useful for information practitioners in the field (see, e.g. Lee *et al.*, 2018; Ozeran and Martin, 2019; Papachristopoulos *et al.*, 2019; Wang, 2022), will educators and librarians designing learning experiences apply such methods in situ? Awareness and training will be key.

Conclusion

Today's young people, more than any previous generation, have a personal stake in their ability to function with data. Future job prospects might hinge on their ability to participate in the new data economy. But equally, young people are themselves the subjects of the data economy, being the most thoroughly measured, tracked and analyzed generation (Taylor and Rooney, 2016). Informal learning frameworks and, particularly, participatory design activities offer promising avenues for engaging youth in data literacy learning. First, the products designed by teens can be more meaningful and authentic to other teens who engage with data literacy activities designed with, for, and by teens. Second, the very act of designing is, itself, a learning experience for teen designers because, to get to the point at which they can recommend, create and critique data activities, they must first gain some foundational knowledge *about* data literacy. This presumes that adults (or peers) with some expertise in data literacy must be a part of the design process, helping to activate this knowledge transfer, which in turn can open up questions about intergenerational interaction. One instrumental benefit of youth engagement with data literacy, especially through co-design, is to support democratic systems, a la John Dewey (1916). Other benefits include advancing personal and societal economic development, increasing community well-being and providing an enhanced appreciation for an intrinsically fascinating area of techno-cultural development (NASEM, 2016). Given all these benefits, it is important to gain a deeper understanding of the nature of the interpersonal dialogue that occurs during the co-design of data literacy activities for teens – a task this study set out to do.

In summary, in collaboration with public libraries and teens, we conducted a project to design, develop and test potential after-school data literacy activities for a teen audience at the public library. This article presents the findings pertaining to the linguistic indicators of the nature of teen thinking, their emotions, level of engagement and the power relationships between teens and adults in the context of conversations about data.

With full awareness of the limitations of computerized language measures and regardless of recent improvements in NLP, these are still probabilistic systems that can ignore the context of a sentence, a word or a local idiom; may not detect irony, satire and metaphor; and are not fully transparent to its users. It is also not clear to what degree that LIWC-22 accounts for teen language. LIWC-22 scores are calculated from the entire pool of words, rather than a specific sentence or paragraph in context. We therefore see NLP techniques as one part of a multimethod package of tools, working side-by-side traditional qualitative methods. In terms of the study of teen–adult interactions in co-design, we recommend a hybrid approach to triangulate the findings, using both machine and human to get at the deeper psychological layers embedded in language.

We hope that this article will offer methodological guidance to fellow researchers and practitioners interested in exploring the potential of NLP as a tool for designing meaningful data literacy education, thus stimulating further research and discussion on the ways to empower youth to engage more actively in informal learning about data.

References

Acker, A. and Bowler, L. (2018), "Youth data literacy: teen perspectives on data created with social media, and mobile device ownership", *Hawaii International Conference on System Sciences 2018 (HICCS)*. January 3-6, 2018, available at: http://hdl.handle.net/10125.50130

Arowosegbe, A. and Oyelade, T. (2023), "Application of natural language processing (NLP) in detecting and preventing suicide ideation: a systematic review", *International Journal of Environmental Research and Public Health*, Vol. 20 No. 2, p. 1514, doi: 10.3390/ijerph20021514.

Bilstrup, K.E.K., Kaspersen, M.H., Lunding, M.S., Schaper, M.M., Van Mechelen, M., Tamashiro, M.A., Smith, R.C., Iversen, O.S. and Petersen, M.G. (2022), "Supporting critical data literacy in K-9 education: three principles for enriching pupils' relationship to data", *IDC '22: Interaction Design and Children*, Braga Portugal June 27-30, 2022, Now Publishers, New York, NY, pp. 225-236.

Bowler, L., Acker, A., Jeng, W. and Chi, Y. (2017), "'It lives all around us': aspects of data literacy in t'en's lives", *Proceedings of the Association for Information Science and Technology*, Vol. 54 No. 1, pp. 27-35.

Bowler, L., Aronofsky, M., Milliken, G. and Acker, A. (2020), "Teen engagements with data in an after-school data literacy programme at the public library", *Proceedings of ISIC, the Information Behaviour Conference*, *Pretoria, South Africa*, 28 September-1 October, 2020, *Information Research*, 25(4), paper isic2015, available at: http://InformationR.net/ir/25-4/isic2020/isic2015.html (Archived by the Internet Archive at https://bit.ly/3a6DBwv), doi: 10.47989/irisic2015.

Bowler, L., Wang, K., Lopatovska, I. and Rosin, M. (2021), "The meaning of 'participation' in co-design with children and youth: relationships, roles, and interactions", *Proceedings of the Association for Information Science and Technology*, Vol. 58 No. 1, pp. 13-24, doi: 10.1002/pra2.432.

Bowler, L., Rosin, M., Lopatovska, I. and Vroom, L. (2022a), "Co-designing data labs at the public library: data literacy with, for, and by teens", *iConference 2022: Information for a Better World: Shaping the Global Future*, February –8 – March 4, 2022, available at: http://hdl.handle.net/2142/113734

Bowler, L., Rosin, M., Lopatovska, I. and Vroom, L. (2022b), "Methods of engaging teens in conversations about personal digital data: public library context", in Filipiak, D. and Kalir, J.H. (Eds), Proceedings of the 2022 Connected Learning Summit, Pittsburgh, PA, ETC Press, July 28-29, 2022, pp. 9-17, available at: https://press.etc.cmu.edu/proceedings/proceedings-2022-connected-learning-summit

Boyd, R.L. and Pennebaker, J.W. (2016), "A way with words: using language for psychological science in the modern era", in Dimofte, C., Haugtvedt, C.P. and Yalch, R.F. (Eds), *Consumer Psychology in a Social Media World*, Routledge/Taylor and Francis Group, Abingdon, UK, pp. 222-236.

Boyd, R.L. and Schwartz, H.A. (2021), "Natural language analysis and the psychology of verbal behavior: the past, present, and future states of the field", *Journal of Language and Social Psychology*, Vol. 40 No. 1, pp. 21-41, doi: 10.1177/0261927X20967028.

Boyd, R.L., Ashokkumar, A., Seraj, S. and Pennebaker, J.W. (2022a), "LIWC-22: descriptive statistics and norms", available at: www.liwc.app

Boyd, R.L., Ashokkumar, A., Seraj, S. and Pennebaker, J.W. (2022b), "The development and psychometric properties of LIWC-22", Austin, TX, University of Texas at Austin, available at: www.liwc.app/static/documents/LIWC-22%20Manual%20-%20Development%20and%20Psychometrics.pdf

Chi, Y., Jeng, W., Acker, A. and Bowler, L. (2018), "Affective, behavioral, and cognitive aspects of teen perspectives on personal data in social media: a model of youth data literacy", *Transforming*

Digital Worlds: 13th International Conference, iConference 2018, Sheffield, UK, March 25-28, 2018, *Proceedings 13*, Springer International Publishing, pp. 442-452.

Cohn, M.A., Mehl, M.R. and Pennebaker, J.W. (2004), "Linguistic markers of psychological change surrounding September 11, 2001", *Psychological Science*, Vol. 15 No. 10, pp. 687-693, doi: 10.1111/j.0956-7976.2004.00741.x.

Csikszentmihalyi, M. (1990), "Flow: the psychology of optimal experience", *Harper Perennial*.

Csikszentmihalyi, M. and Kleiber, D.A. (1991), "Leisure and self-actualization", in Driver, B.L., Brown, P.J. and Peterson, G.L. (Eds), *Benefits of Leisure*, Venture Publications, pp. 91-102.

Csikszentmihalyi, M. and Larson, R. (1984), "Being adolescent: conflict and growing in the teenage years", *Basic Books*.

Deahl, E. (2014), "Better the data you know: developing youth data literacy in schools and informal learning environments", *Available at SSRN 2445621*.

Dewey, J. (1916), *Democracy and Education: An Introduction to the Philosophy of Education*, Macmillan Publishing, New York.

Diaz-Asper, C., Chandler, C., Turner, R.S., Reynolds, B. and Elvevåg, B. (2022), "Increasing access to cognitive screening in the elderly: applying natural language processing methods to speech collected over the telephone", *Cortex*, Vol. 156, pp. 26-38, doi: 10.1016/j.cortex.2022.08.005.

DiSalvo, C., Clement, A. and Pipek, V. (2012), "Participatory design for, with, and by communities", in Simonsen, J. and Robertson, T. (Eds), *Routledge International Handbook of Participatory Design*, Routledge, New York, pp. 182-209.

Druin, A. (1999), "Cooperative inquiry: developing new technologies for children with children", *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, May 1999, *Pittsburgh, PA*, pp. 592-599.

Druin, A. (2002), "The role of children in the design of new technology", *Behaviour and Information Technology*, Vol. 21 No. 1, pp. 1-25, doi: 10.1080/01449290110108659.

Durlak, J.A. and Weissberg, R.P. (2007), "The impact of after-school programs that promote personal and social skills", *Chicago, IL: Collaborative for Academic, Social, and Emotional Learning*, available at: www.casel.org

Falk, J.H. (2001), *Free-Choice Science Education: How we Learn Science outside of School*, Teachers College Press, New York, NY.

Falk, J.H. and Dierking, L.D. (2000), *Learning from Museums: Visitor Experiences and the Making of Meaning*, AltaMira Press, Walnut Creek, CA.

Finzer, W. (2013), "The data science education dilemma", *Technology Innovations in Statistics Education*, Vol. 7 No. 2, doi: 10.5070/T572013891, available at: https://escholarship.org/uc/item/7gv0q9dc

Gray, J., Gerlitz, C. and Bounegru, L. (2018), "Data infrastructure literacy", *Big Data and Society*, Vol. 5 No. 2, p. 2053951718786316.

Hart, R.A. (1992), "Children's participation: from tokenism to citizenship", *UNICEF International Child Development Centre*.

Hidi, S. (1990), "Interest and its contribution as a mental resource for learning", *Review of Educational Research*, Vol. 60 No. 4, pp. 549-571.

Interpreting the LIWC Output (2022), "Introducing LIWC-22: a new set of text analysis tools at your fingertips", available at: www.liwc.app/help/liwc#Interpreting-Output

Jordan, K.N., Sterling, J., Pennebaker, J.W. and Boyd, R.L. (2019), "Examining long-term trends in politics and culture through language of political leaders and cultural institutions", *Proceedings of the National Academy of Sciences*, Vol. 116 No. 9, pp. 3476-3481, available at: www.pnas.org/doi/full/10.1073/pnas.1811987116

Large, A., Nesset, V., Beheshti, J. and Bowler, L. (2006), "'Bonded design': a novel approach to intergenerational information technology design", *Library and Information Science Research*, Vol. 28 No. 1, pp. 64-82, doi: 10.1016/j.lisr.2005.11.014.

Larson, R.W. (2000), "'Toward a psychology of positive youth development", *American Psychologist*, Vol. 55 No. 1, pp. 170-183, doi: 10.1037/0003-066X.55.1.170.

Lee, V.R. (2021), "Youth engagement during making: using electrodermal activity data and first-person video to generate evidence-based conjectures", *Information and Learning Sciences*, Vol. 122 Nos 3/4, pp. 270-291, doi: 10.1108/ILS-08-2020-0178.

Lee, S.P., Perez, M.R., Worsley, M.B. and Burgess, B.D. (2018), "Utilizing natural language processing (NLP) to evaluate engagement in project-based learning", *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, IEEE, pp. 1146-1149.

Leung, Y.W., Wouterloot, E., Wong, J., Gancarz, M., Gratzer, D., Esplen, M.J., Bender, J.L., Adikari, A., de Silva, D., Alahakoon, D. and Hirst, G. (2021), "Natural language processing-based virtual cofacilitator for online cancer support groups: protocol for an algorithm development and validation study", *JMIR Research Protocols*, Vol. 10 No. 1, p. e21453, doi: 10.2196/21453.

LIWC (2022), "Understanding the LIWC summary measures", *Introducing LIWC-22: A new set of text analysis tools at your fingertips*, available at: www.liwc.app/help/liwc#Summary-Measures

LIWC Analysis (2022), "Introducing LIWC-22: a new set of text analysis tools at your fingertips", available at: www.liwc.app/help/liwc#Interpreting-Output

Ma, L., Ou, W. and Sian Lee, C. (2022), "Investigating consumers' cognitive, emotional, and behavioral engagement in social media brand pages: a natural language processing approach", *Electronic Commerce Research and Applications*, Vol. 54, p. 101179, doi: 10.1016/j.elerap.2022.101179.

Matheson, L.A., Wood, E.H. and Franklin, S.V. (2017), "Guided and unguided student reflections", *PERC Proceedings*, *Cincinnati, OH*, July 26-27, 2017, in Ding, L., Traxler, A. and Cao, Y. (Eds), doi: 10.1119/perc.2017.pr.061.

National Academies of Sciences, Engineering, and Medicine (NASEM) (2016), *Science Literacy: Concepts, Contexts, and Consequences*, The National Academies Press, Washington, DC, doi: 10.17226/23595.

National Research Council (2009), "Division of behavioral and social sciences and education", in Bell, P., Lewenstein, B., Shouse, A.W. and Feder, M.A. (Eds), *Learning Science in Informal Environments: People, Places, and Pursuits*, Committee on Learning Science in Informal Environments, Board on Science Education, Center for Education, The National Academies Press, Washington, DC.

Oliver, K.M., Houchins, J.K., Moore, R.L. and Wang, C. (2021), "Informing makerspace outcomes through a linguistic analysis of written and video-recorded project assessments", *International Journal of Science and Mathematics Education*, Vol. 19 No. 2, pp. 333-354.

Ozeran, M. and Martin, P. (2019), "Good night, good day, good luck", *Information Technology and Libraries*, Vol. 38 No. 2, pp. 49-57, doi: 10.6017/ital.v38i2.10921.

Papachristopoulos, L., Ampatzoglou, P., Seferli, I., Zafeiropoulou, A. and Petasis, G. (2019), "Introducing sentiment analysis for the evaluation of library's services effectiveness", *Qualitative and Quantitative Methods in Libraries*, Vol. 8 No. 1, pp. 99-110.

Ryu, S. and Lombardi, D. (2015), "Coding classroom interactions for collective and individual engagement", *Educational Psychologist*, Vol. 50 No. 1, pp. 70-83.

Silvia, P.J. (2008), "Intere – t – The curious emotion", *Current Directions in Psychological Science*, Vol. 17 No. 1, pp. 57-60.

Sowmya, V.B., Majumder, B., Gupta, A. and Surana, H. (2020), "Practical natural language processing: a comprehensive guide to building real-world NLP systems (First edition)", *O'Reilly Media*.

Speer, A.B., Perrotta, J., Tenbrink, A.P., Wegmeyer, L.J., Delacruz, A.Y. and Bowker, J. (2022), "Turning words into numbers: assessing work attitudes using natural language processing", *Journal of Applied Psychology*, Vol. 108 No. 6, pp. 1027-1045.

Subramaniam, M. (2016), "Designing the library of the future for and with teens: librarians as the 'connector' in connected learning", *The Journal of Research on Libraries and Young Adults*, Vol. 7 No. 2, pp. 1-18.

Tanana, M.J., Soma, C.S., Kuo, P.B., Bertagnolli, N.M., Dembe, A., Pace, B.T., Srikumar, V., Atkins, D.C. and Imel, Z.E. (2021), "How do you feel? Using natural language processing to automatically rate emotion in psychotherapy", *Behavior Research Methods*, Vol. 53 No. 5, pp. 2069-2082, doi: 10.3758/s13428-020-01531-z.

Taylor, E. and Rooney, T. (2016), "Digital playgrounds: growing up in the surveillance age", *Surveillance Futures Social and Ethical Implications of New Technologies for Children and Young People (Emerging Technologies, Ethics, and International Affairs)*, 1st ed. Routledge, London; New York, NY, pp. 13-28.

Tygel, A. and Kirsch, R. (2015), "Contributions of Paulo Freire for a critical data literacy", Proceedings of Web Science 2015 Workshop on Data Literacy, pp. 318-334.

Walker, J., Marczak, M., Blyth, D. and Borden, L. (2005), "Designing youth development programs: toward a theory of developmental intentionality", in Mahoney, J.L., Larson, R.W. and Eccles, J.S. (Eds), *Organized Activities as Contexts of Development*, Lawrence Erlbaum, Mahwah, NJ, pp. 399-418.

Wang, Y. (2022), "Using machine learning and natural language processing to analyze library chat reference transcripts", *Information Technology and Libraries*, Vol. 41 No. 3, pp. 1-10, doi: 10.6017/ital.v41i3.14967.

Yip, J.C., Lee, K.J. and Lee, J.H. (2019), "Design partnerships for participatory librarianship: a conceptual model for understanding librarians co designing with digital youth", *Journal of the Association for Information Science and Technology*, Vol. 71 No. 10, pp. 1242-1256, doi: 10.1002/asi.24320.

Yip, J.C., Sobel, K., Pitt, C., Lee, K.J., Chen, S., Nasu, K. and Pina, L.R. (2017), "Examining adult-child interactions in intergenerational participatory design", *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, May 2017, *Denver, CO*, pp. 5742-5754.

Further reading

Boyd, R.L. (2017), "Psychological text analysis in the digital humanities", in Hai-Jew, S. (Ed.), *Data Analytics in Digital Humanities*, Springer International Publishing, New York, pp. 161-189, doi: 10.1007/978-3-319-54499-1_7.

LIWC Home (2023), "Introducing LIWC-22: a new set of text analysis tools at your fingertips", available at: www.liwc.app/

Pennebaker, J.W. (2011), *The Secret Life of Pronouns: What Our Words Say about Us*, Bloomsbury Press, New York.

Sharma, A., Barrett, K. and Stapelfeldt, K. (2022), "Natural language processing for virtual reference analysis", *Evidence Based Library and Information Practice*, Vol. 17 No. 1, pp. 78-93, doi: 10.18438/eblip30014.

Smith-Keiling, B.L. and Hyun, H.I.F. (2019), "Applying a computer-assisted tool for semantic analysis of writing: uses for STEM and ELL", *Journal of Microbiology and Biology Education*, Vol. 20 No. 1, p. 70.

Sowmya, V.B., Majumder, B., Gupta, A. and Surana, H. (2020), *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*, 1st ed., O'Reilly Media, Sebastopol, CA.

Tausczik, Y.R. and Pennebaker, J.W. (2010), "The psychological meaning of words: LIWC and computerized text analysis methods", *Journal of Language and Social Psychology*, Vol. 29 No. 1, pp. 24-54, doi: 10.1177/0261927X09351676.

About the authors

Leanne Bowler is a Professor at the School of Information, Pratt Institute, New York City. Her research and teaching focuses on young peoples' interactions with information and data, their technology practices, STEM learning and how family, teachers and out-of-school organizations such as libraries and museums can support young people's competencies in a sociotechnical world. Leanne Bowler is the corresponding author and can be contacted at: lbowler@pratt.edu

Irene Lopatovska is a Professor at the School of Information, Pratt Institute, New York City. Her research interests include human information interaction and affective computing. Her current research examines interactions with conversational agents like Amazon's Alexa. Dr Lopatovska is experienced in conducting user studies in various settings and with different populations, including a collaboration with the Brooklyn Public Library to examine children's literacy skills.

Mark Rosin is an Associate Professor in the Department of Math and Science at Pratt Institute in New York City and Executive Director at Guerilla Science, an international science and society organization. His work focuses on applied mathematics; computer algorithms; the connections between science, art and culture; and their realizations in the public places and spaces that science is least expected. He is the winner of the Early Career Award for Public Engagement with Science from the American Association for the Advancement of Science and an Ashoka Foundation Emerging Innovator. His work with Guerilla Science has been featured in *The New Yorker*, *The New York Times* and *VICE*.