Computer Science Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncse20

Investigating relationships of sentiments, emotions, and performance in professional development K-12 CS teachers

Yi Liu, Leen-Kiat Soh, Guy Trainin, Gwen Nugent & Wendy M. Smith

To cite this article: Yi Liu, Leen-Kiat Soh, Guy Trainin, Gwen Nugent & Wendy M. Smith (16 Jan 2024): Investigating relationships of sentiments, emotions, and performance in professional development K-12 CS teachers, Computer Science Education, DOI: 10.1080/08993408.2023.2298162

To link to this article: https://doi.org/10.1080/08993408.2023.2298162

Published online: 16 Jan 2024.

COMPUTER SCIENCE EDUCATION
https://doi.org/10.1080/08993408.2023.2298162

Routledge
Taylor & Francis Group

Check for updates

Investigating relationships of sentiments, emotions, and performance in professional development K-12 CS teachers

Yi Liu[a], Leen-Kiat Soh[a], Guy Trainin[b], Gwen Nugent[c] and Wendy M. Smith[d]

[a]School of Computing, University of Nebraska-Lincoln, Lincoln, USA; [b]Department of Teaching, Learning, and Teacher Education, University of Nebraska-Lincoln, Lincoln, USA; [c]Nebraska Center for Research on Children, Youth, Families and Schools, University of Nebraska-Lincoln, Lincoln, USA; [d]Center for Science, Mathematics, and Computer Education; Department of Mathematics, University of Nebraska-Lincoln, Lincoln, USA

ABSTRACT

Background and Context: Professional development (PD) programs for K-12 computer science teachers use surveys to measure teachers' knowledge and attitudes while recognizing daily sentiment and emotion changes can be crucial for providing timely teacher support.

Objective: We investigate approaches to compute sentiment and emotion scores automatically and identify associations between the scores and teachers' performance.

Method: We compute the scores from teachers' assignments using a machine-assisted tool and measure score changes with standard deviation and linear regression slopes. Further, we compare the scores to teachers' performance and post-PD qualitative survey results.

Findings: We find significant associations between teachers' sentiment and emotion scores and their performance across demographics. Additionally, we find significant associations that are not captured by post-PD qualitative surveys.

Implications: The sentiment and emotion scores can viably reflect teachers' performance and enrich our understanding of teachers' learning behaviors. Further, the sentiment and emotion scores can complement conventional surveys with additional insights related to teachers' learning performance.

ARTICLE HISTORY
Received 20 June 2023
Accepted 19 December 2023

KEYWORDS
Professional development; K-12 teachers; computer science education; sentiment analysis; emotion analysis; machine-based text analysis

1. Introduction

With the advent of the Computer Science (CS) For All initiative (Smith, 2016) in the US and the Informatics For All initiative in Europe (Informatics Europe, 2016), we have seen significant demand for professional development (PD) in CS for K-12 teachers (e.g. Menekse, 2015). Studies have reported on different PD designs and their effectiveness. In recent years, many K-12 schools have started incorporating CS curricula to meet the need to introduce CS to K-12 students. As many schools lack resources to create PD to support educators (Kafai et al., 2020; Stange, 2020), more professional support is needed for CS teachers (Falkner et al., 2018; Ni et al., 2023).

CONTACT Yi Liu yil@cse.unl.edu School of Computing, University of Nebraska-Lincoln, Lincoln, USA

Furthermore, because teachers who are committed to teaching CS, often have limited exposure to CS, introducing and exposing CS teaching concepts, tools, and resources for the classroom is an important aspect of PD (Broneak & Rosato, 2021; Jocius et al., 2020; Kaya et al., 2021), in addition to addressing teachers' needs, concerns, and challenges such as curricular resources, community or peer support, and institutional support (e.g. Allison, 2023; Ravitz et al., 2017; Sadik & Ottenbreit-Leftwich, 2023; Yadav et al., 2016). A computer programming language can be challenging for teachers without a CS background due to its syntax and programming environment, and thus intimidating to teachers new to CS education. For example, statements and keywords in programming are simplified for convenient usage by developers. However, this simplification could create initial confusion for teachers and create a resistance to further learning. To make such teaching easier for students and teacher developers have created teaching tools and resources including visual programming tools (e.g. Scratch and Blockly), educational hardware (e.g. Micro:bit), and online coding platforms (e.g. CodeHS, code.org). While there are PDs that used block-based and visual-based programming instruments that avoid the challenges of text-based programming (Faherty et al., 2020; Siever & Rogers, 2019), there are still demands for K-12 teachers learning text-based programming as there is a need for transitioning from block-based and visual-based programming to text-based programming (e.g. for high school students preparing for higher education) (Dawson, 2021). Text-based programming languages are different from block-based and visual-based programming languages. For example, the text-based programming language has programming grammar (or syntax) that students must follow to allow the code to run correctly, while block-based and visual-based programming languages hide the challenges in programming grammar to let students focus on program logic. However downstream expectations in post-secondary education emphasize text-based programming languages such as C++, Python, and Java. Thus, K-12 teachers need to help students who are interested in studying CS in college to complete the transition from block- and visual-based programming to text-based programming.

PD providers often measure teachers' engagement and changing attitudes, so they can better support teachers and help mitigate negative responses to the challenges of learning CS and programming. Emotions and attitudes have been reported to play important role in students' perceptions of learning programming (Bosch & D'Mello, 2017; Lishinski et al., 2017; Malmi et al., 2020; Ruiz et al., 2016). It is likely that teachers experience similar emotional and sentimental effects when they learn to program during PD. Sentiments (negative or positive) are general attitudes toward PD, while emotions (e.g. joy, anger, and fear) are attitudes of specific aspects toward PD. For example, Negative sentiments or emotions (e.g. numerous complaints) from a teacher could signal frustration and disengagement of the teacher, leading to ineffective learning outcomes. On the other hand, a strong positive sentiment or emotion towards a subject – such as excitement about learning new topics – could indicate that teachers welcome the challenge and are more likely to persist in learning. Thus, by gauging teachers' sentiments and emotions, organizers could use teachers' motivation and engagement to adjust the PD contents and scaffolding responsively.

The traditional mode of gauging sentiment and emotion are surveys. Surveys present challenges to continuously gauge the variability in teachers' sentiments and emotions with validity. First, it is not practical to administer surveys daily as that would lead to survey fatigue (Porter et al., 2004), which reduces the validity of facing challenges during coursework. Also, survey responses suffer from desirability bias in which teachers might react in ways they think the researcher is expecting them to (Fisher, 1993; Rossi et al., 2013).

Assignment writings can provide an alternative way to measure teacher sentiments and emotions during CS PD. These writings refer to assignments that allow teachers to express their subjective opinions, not programming assignments. These assignments include reflections about the PD and discussion boards. Hence, we could derive teachers' sentiments and emotions from the text expressing subjective opinions. Since the assignments are synchronized with the PD assignment, it does not create an additional burden for the teachers. At the same time, the assignment is part of the course work which will accurately reflect the actual mindset of teachers while they work on assignments. Unlike surveys, authentic assignments allow a valid and potentially rich data from which to gauge a teacher's sentiment and emotion during PD.

We are motivated to investigate and develop an effective method for gauging teachers' sentiments and emotions to facilitate timely intervention and support for teachers' needs during PD. Specifically, as discussed above, teachers' sentiments and emotions reflect their perceptions of the subject matter and their learning of the subject matter. By gauging teachers' sentiments and emotions, one can gain additional insights into teachers' learning behaviours and performance during PD. However, using conventional methods such as surveys (e.g. daily assessments of teachers' attitudes or sentiments about the PD content covered on each day) may be impractical because of potential survey fatigue (Porter et al., 2004). As a result, we explore using what the teachers are already required to complete each day, namely, their daily assignment writings, to gauge their sentiments and emotions about what they learn. Our investigation is also driven by the following research question: *Can analyses based on teachers' assignment writings provide a timely mode of gauging teachers' sentiments and emotions related to their learning during CS PD?*

In this paper, we investigated associations between sentiment/emotion scores and teacher performance in pre- and post-PD scores using backward linear regression (Field, 2013; Montgomery et al., 2021). We report on our analysis of teachers' data collected from four cohorts of summer CS PD institutes (Nugent et al., 2020). We collect texts from teachers' writings to the assignments to compute numerical scores representing teachers' sentiments and emotions using NRCLex (Mohammad & Turney, 2013). NRCLex has been widely used in sentiment and emotion analyses related to research and project (Dolianiti et al., 2019; Shaik et al., 2023).

Further, we also perform backward linear regression analyses between the teachers' responses to conventional post-PD surveys (Nugent et al., 2020) and their performance in their summer PD to investigate if there are additive perspectives provided by the sentiment/emotion analysis. Specifically, the post-PD survey has six numerical outcomes: (1) CS self-efficacy, (2) CS content, (3) CS pedagogy, (4) CS attitudes, (5) personal interest, and (6) perceived value. We computed the association between these six outcomes and teachers' performance. Then, we compare that analysis with the perspective provided by the sentiment/emotion analysis. We found that certain perspectives provided by the sentiment/emotion analysis added to the conventional post-PD survey to complement the conventional survey approach towards better responding to challenges, changes, or frustration that teachers might encounter during a PD in order to help them succeed in the PD.

Here we summarize the findings of the above investigations.

(1) *Analysing sentiments and emotions derived from teachers' submitted assignments across multiple days in order to, prompt intervention or inquiry to address potential concerns of teachers is viable during a PD (Sections 4.2 and 4.3)*. By using both analyses, at least three sentiment or emotion scores can be used as an indicator of the challenges, changes, or frustrations that might impact teachers' performance for different demographic groups of teachers (Sections 4.2.2 and 4.3.2). Further, we see evidence that supports the viability of the automated text-based sentiment/emotion analysis tool – NRCLex (Section 4.3).
(2) *The sentiment and emotion analyses based on teachers' assignments provide additional insights and perspectives for understanding the mediating role of sentiments and emotions on teachers' performance during a PD (Section 4.4)*. For example, we see significant relationships between teachers' sentiment and emotions scores and their performance that are not found based on conventional post-PD surveys (Section 4.4.2).
(3) *The variability in and the trend of sentiments or emotions can be used to monitor changes in teachers during PD to monitor teachers' performance (Sections 4.2.2 and 4.3.2)*. We observe that the sentiment and emotion scores can supplement a conventional post-PD survey in modelling both teachers' post-PD performance and incremental performance in providing real-time feedback (Section 4.4).
(4) *Using the variability in and trend of sentiments or emotions provides additional nuances to simply using the overall sentiment/emotion (Sections 4.2 and 4.3)*. Specifically, the variability in and trend of sentiments or emotions revealed associations between the sentiment or emotion and teachers' performance for city/suburban teachers, and middle school teachers, while there was no association found using only the overall sentiment/emotion (Sections 4.2.2 and 4.3.2).

In section 2 we provide an overview of related works. Section 3 describes the methods and algorithms used in our investigation. Section 4 gives the analyses of three investigations and reports on the results. Section 5 concludes and presents future work.

2. Background and related work

2.1. CS professional development

PD is an effective way to help teachers to get access to teaching tools and resources that can be used in their classrooms and to build their pedagogical and content knowledge, and confidence to deliver CS instructions. There are teaching tools and resources to mitigate the challenges of advanced programming to allow high school students to learn and understand advanced CS topics. Hjorth (2021) proposed the NLP4All to bring natural language processing into high school classes. NLP4All allows a non-programmer to explore the text classification based on natural language processing, where the text is the post on tweets. Siever and Rogers (2019) propose to introduce CS topics related to robotics, the Internet of Things, and wireless communication using a cheap and entry-level platform named Micro:bit. Biswas et al. (2019) propose a web-based learning environment, namely C2STEM, to allow students to develop computational skills using CT activities in realistic scenarios. Further, the CS concepts, such as variables, conditionals, loops, arrays, and functions, and the underlying computational thinking (CT) are CS content knowledge needed for CS education (Grover et al., 2020). Incidentally, it has also been reported that introducing CT into education courses can effectively influence preservice teachers' understanding of CS concepts (Yadav et al., 2014).

Improving teachers' self-efficacy in teaching CS during PD is important, because their self-efficacy can predict both teaching behaviours and student outcomes (Zhou et al., 2020). Hamlen Mansour et al. (2023) showed that student gains were higher for those whose teachers were more confident in their ability to teach CS. There, by showing teachers how to get accessible resources such as animated videos introducing CS concepts, teachers can gain confidence to independently find necessary materials to create a curriculum for their students without getting lost in the vast Internet. Furthermore, PD that uses the teacher-learner-observer (TLO) model is effective to develop teachers' confidence (Loucks-Horsley et al., 2009). Margolis et al. (2017) introduced the TLO model where teachers learn content in the context of teaching lessons, observing, and reflecting to feature a creative, active, and participatory PD. In particular, PD in text-based programming serves to encourage and facilitate teachers in getting involved in CS education and building their self-efficacy (Bandura & Watts, 1996) in teaching CS.

Due to different resources and types of support available, CS teachers who are from different regions, such as rural, suburban, and urban regions (Ryoo et al., 2021), and have different backgrounds (Broneak & Rosato, 2021; Jocius et al., 2020; Kaya et al., 2021) have different access to teaching knowledge and curricular materials. PD can bring CS content to teachers from different regions and different backgrounds. Sentance and Humphreys (2018) report the importance of a community of practice as part of a PD. Similarly, through their comprehensive review of K-12 CS PDs, Ni et al. (2023) also indicated the need for professional learning communities (PLCs) to scale up PD and sustain teaching capacity. Sauppé et al. (2019) present the experience of collaborating with K-12 teachers in low-population regions to increase the comfortableness of teachers to introduce CS into their classrooms. Mouza et al. (2023) reported partnering undergraduates and K-12 school teachers in their PD design and found that undergraduates were able to connect knowledge of computing to pedagogy and technology to assist teachers in the implementation of CS instruction. These studies have implications for supporting K-12 teachers for sustainable CS teaching.

2.2. Sentiment and emotion analysis

Emotions are a series of coordinated psychological systems including affective, cognitive, motivational, expressive, and peripheral physiological processes (Damasio, 2004). Affective processes are assumed to be central to emotions, and to be physiologically bound to subsystems of the limbic system (Fellous & Ledoux, 2005). For example, joy includes components such as positive feelings (affective component), senses of well-being (cognitive), positive (motivational), passive facial expression (expressive), and even senses of tiredness (physiological). Achievement emotions are defined as "emotions tied directly to achievement activities or achievement outcomes" (Pekrun, 2006). Studies of achievement emotions typically focus on emotions relating to achievement outcomes and achievement-related activities. Examples include the enjoyment of learning, boredom experienced in the classroom, and frustration with difficult tasks or negative outcomes such as a low grade. Pekrun and his colleagues separated achievement emotions into activity emotions and outcome emotions pertaining to the outcomes of these activities including anticipatory emotions (e.g. hope for success, anxiety of failure (Pekrun et al., 2002, 2009)).

Achievement emotions can be momentary occurrences (state achievement emotions) in reaction to a given situation and time (e.g. frustration at a difficult computer science task). They can also be as habitual, recurring emotions typically experienced by an individual for example joy of interacting with exercises on a familiar platform such as Scratch. Trait achievement emotions are separated from state achievement emotions by their dependence on variation across time.

For computer science education, it has been found that sentiment and emotion play a significant role in affecting students' learning outcomes. Malmi et al. (2020) summarized substantial theoretical development addressing relationships between learning programming and students' emotions, attitudes, and self-efficacy. They found the importance of these factors and the need for tools capturing these factors in supporting programming education. This finding also has been supported by various studies such as Law et al. (2010), Anastasiadou and Karakos (2011), and Kinnunen and Simon (2012), which collectively suggest that positive attitudes, confidence, and the reduction of negative emotions like computer anxiety are essential for effective learning in programming. Further, emotional factors have been investigated to explore their impacts on learning outcomes. For example, Kuo et al. (2013) suggested investigating emotional effects that could improve students' self-efficacy for better learning performance. Bosch and D'Mello (2017) found that two combinations of emotions (confusion & frustration and curiosity & engagement) dominated the impact on the learning outcome. Lishinski et al. (2017) identified that emotions such as frustration and self-efficacy affected students' learning outcomes in both the short and long terms. These findings advocate for the integration of sentiment-aware and emotion-aware technologies for effective computer science education, such as EarSketch (Magerko et al., 2016) and adaptive web-based learning environments (Cabada et al., 2018; Liao et al., 2019).

On the other hand, capturing sentiment and emotion has been challenging in practice for a class. Conventional approaches capture the data using questionnaires and surveys (e.g. Ruiz et al., 2016; Scott & Ghinea, 2014). However, these approaches require additional work for the student, which can lead to survey fatigue (Porter et al., 2004). This suggests the necessity for innovative approaches that can efficiently capture sentiment and emotion metrics without impacting the student experience.

With the development of natural language processing, text-based analysis approaches have made sentiment and emotion capturing possible without asking for additional work. These approaches provide tools allowing us to evaluate sentiments and emotions based on students' assignment writings during the class. Valence Aware Dictionary for Sentiment Reasoning (VADER) (Hutto & Gilbert, 2014) analyses polarity, positive or negative, and intensity of sentiment using the lexicon and rule-based sentiment analysis. TextBlob (Loria et al., n.d.) is a Python library that provides textual data processing for natural language processing tasks. TextBlob provides rule-based sentiment analysis that evaluates the polarity and the intensity of sentiment. Both transformer and recurrent neural network models are based on deep learning and provided by Flair (Akbik et al., 2019), a state-of-the-art natural language processing library. They evaluate the polarity of sentiment and show the confidence level of the prediction of the polarity. NRCLex measures emotional scores from the body of texts using a lexicon dictionary of words (Mohammad & Turney, 2010). The lexicon dictionary is constructed using crowdsourcing (Mohammad & Turney, 2013). When comparing these tools (VADER, TextBlob, Flair, and NRCLex), it becomes evident that NRCLex stands out because it offers the most comprehensive analysis results. These results include the relative intensity of eight emotion factors (anticipation, joy, trust, surprise, disgust, fear, anger, and sadness) and two sentiment polarities (positive and negative).

3. Methodology

In the following section, we discuss the methods for obtaining sentiment and emotion scores and teachers' performance scores and carrying out the association evaluations based on backward regression. Section 3.1 discusses the method for sentiment and emotion analyses. Section 3.2 discusses the computation to obtain teachers' performance scores. Section 3.3 describes the association evaluation in our investigation. Section 3.4 shows the Professional Development (PD) design and characteristics. Finally, Section 3.5 describes the data collection in our investigation.

3.1. Sentiment and emotion analyses

This investigation evaluated teachers' sentiments based on textual writings of teachers collected as part of their assignments during the 2-week summer institute using text-based sentiment analysis algorithms. Two sentiment scores and eight emotion scores are evaluated using NRCLex (Mohammad & Turney, 2010). The sentiment score is a pair of polarity scores: positive and negative sentiments. The eight emotion scores represent eight emotion effects: (1) anticipation, (2) joy, (3) trust, (4) surprise, (5) disgust, (6) fear, (7) anger, and (8) sadness. NRCLex, implemented in Python, evaluates sentiment and emotion scores based on a word-wise analysis using term-sentiment and term-emotion association lexicons. The term-sentiment and term-emotion association lexicons are based on the National Research Council Canada affect lexicon and the NLTK library's WordNet (Princeton University, n.d.) synonym sets, which contain the sentiment and emotion scores for approximately 27,000 words. The NRCLex tool outputs ten scores (two sentiments and eight emotions) representing the intensity of the sentiments and emotions of given texts. First, NRCLex looks up the sentiment and emotion score for each word in the given texts. Then, statistics, thresholds, and rules are applied to these scores of all words to determine the sentiment and emotion scores of the given texts. For further details on the innerworkings of NRCLex, interested readers are referred to Mohammad (2021, 2022).

Further, we compute three metrics: (1) overall sentiment and emotion score, (2) variability in sentiment or emotion scores of the teachers, and (3) trend in sentiment or emotion scores of the teachers over time. To compute the overall sentiment or emotion score of a teacher, we combine all the assignment writings that the teacher submitted during the PD into a single file. Then, we apply NRCLex to compute a single score for that teacher. To compute the variability, on the other hand, we compute the sentiment or emotion score of each assignment submitted by the teacher using NRCLex, and then compute the standard deviation of the scores. Since it is possible that how a sentiment trends up or down can reflect changes in a teacher's emotion, we further compute the trend of a teacher's sentiment or emotion scores. More specifically, we first order teachers' assignment writings based on the due dates of the assignments. We then compute the sentiment score of each assignment and then compute the slope of the scores for each instructor using linear regression.

3.2. Performance analysis

Teacher performance is measured based on their knowledge of computer science using pre- and post-tests that have been previously validated in beginning undergraduate CS courses. The performance assessment consists of two parts: (1) teachers' knowledge of CS concepts (e.g. selection statements, functions, and sorting) (Shell & Soh, 2013) and (2) teachers' computational thinking (CTCAST: Peteranetz et al., 2020). Based on these two parts, an aggregated score based on the pre- and post-tests is computed to indicate teacher performance for the summer institute.

More specifically, teacher performance is based on teachers' pre-and post-test scores and has two parts, CS concepts scores (CS scores) and CT and pedagogical concepts scores (CT scores) collected at the end of the summer institute. We compute a compound final score for each teacher based on both parts of the scores. Specifically, the compound final grade of each teacher, t, is a summation of two score parts normalized by the summation of maximum scores of the two parts using the following equation:

$$Grade_t^{post} = \frac{S_{ct}^{t,post} + S_{cs}^{t,post}}{S_{ct}^{max} + S_{cs}^{max}} \tag{1}$$

where $S_{ct}^{t,post}$ is CT scores of the final grade for the corresponding teacher, t, and $S_{cs}^{t,post}$ is CS scores of the final grade for the corresponding teacher, t. Note that, the maximum score of the CT and pedagogical concepts score, S_{ct}^{max}, is 18, and the maximum score of the CS concepts score, S_{cs}^{max}, is 13.

Further, we compute the percentage change of performance that is based on the difference between the final grade at the end of a summer institute and the pre-test grade at the beginning of the institute. The percentage change of performance is positive if the final grade is higher than the pre-test grade, otherwise it is negative. Specifically, the compound percentage change of performance (incremental performance) of each teacher, t, is the final grade, $Grade_t^{post}$ subtracted by the pre-test grade, $Grade_t^{pre}$:

$$Grade_t^{pre} = \frac{S_{ct}^{t,pre} + S_{cs}^{t,pre}}{S_{ct}^{max} + S_{cs}^{max}} \tag{2}$$

$$\Delta Grade_t = Grade_t^{post} - Grade_t^{pre} \tag{3}$$

In our investigation, we employ the compound performance score ($Grade_t^{post}$ and $\Delta Grade_t$) based on CS scores and CT scores. Note that we combine both the CS and CT scores to assess teachers' performance in our investigation. Our reasons are as follows. First, the PD courses integrate the components of CS, CT, and pedagogical concepts comprehensively (Section 3.4). The assignments submitted by the teachers reflect on their learning and understanding processes of these concepts as a whole. Thus, it is not necessary to use the CS or CT scores separately to serve as teachers' performance. Second, we did not observe statistically significant differences between using CS or CT scores separately and using the compound CS and CT scores. Specifically, we computed the difference between each pair of the three sets of correlation coefficients between teachers' performance and their sentiment and emotion scores using: (1) only CS scores, (2) only CT scores, and (3) the compound CS and CT scores. Using student t-test, all differences yielded p-values > 0.05.

3.3. Backward regression analysis

Backward regression is a statistical technique that is commonly used in research to identify the variables that have the strongest relationship with a dependent variable. This technique is often preferred over forward regression because it allows researchers to start with a model that includes all of the variables of interest and then gradually removes variables that are found to have a weak relationship with the dependent variable (Cohen et al., 2013). This results in a more parsimonious model that includes only the variables that are most important for predicting the outcome variable.

Backward regression is particularly useful in situations where the number of independent variables is large and where there is a risk of overfitting the model. Overfitting occurs when a model is too complex and includes variables that do not actually contribute to predicting the dependent variable (Field, 2013). Backward regression helps to avoid overfitting by gradually removing variables that do not contribute to the model, resulting in a more accurate and interpretable model.

Another advantage of backward regression is that it provides a systematic approach to variable selection, which is important for ensuring the reliability and validity of the results. By starting with a full model that includes all of the variables of interest, researchers can be sure that they have not overlooked any potential predictors of the dependent variable (Tabachnick & Fidell, 2013).

In summary, backward regression is a useful statistical technique for identifying the most important predictors of a dependent variable in a large dataset. It allows researchers to systematically remove variables that do not contribute to the model, resulting in a more accurate and interpretable model that is less likely to overfit the data.

3.4. Summer institutes

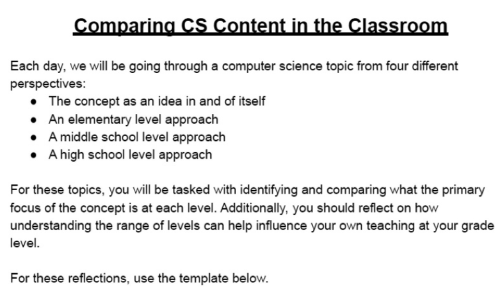
We investigated four cohorts of PD each going through a 2-week summer institute. All cohorts were introduced to two types of content: (1) CS concepts and (2) Computational Thinking (CT) and pedagogical concepts for K-12 CS school teachers. CS concepts included an introduction to concepts and programming skills. CT and pedagogical concepts include an introduction to teaching resources and pedagogical strategies for K-8 CS instruction. In the cohorts, the instructional team consisted of K-12 CS school teachers, assisted by a university CS professor and teaching assistants who were university graduate students from the CS department.

The summer institute for the cohort covers two components: computer programming and CS education pedagogy. First, it familiarizes teachers with problem-solving approaches using computer programming. Specifically, during the summer institute, teachers gained hands-on experience in computer programming to understand fundamental computational concepts, such as variables, conditions, loops, and functions. This practical experience also enhanced teachers' ability to convey these concepts to their students more effectively. Second, the summer institute introduced pedagogical knowledge related to teaching tools for CS education. For example, teachers discussed the benefits of using pedagogical strategies based on available teaching tools such as Ozobots and Micro:bit in their classes. This knowledge enriched their teaching method enabling them to design effective instruction, and assessment for effective CS education delivery.

During each 2-week summer institute, daily assignments required teachers to reflect on the day's learning including CS, CT, and pedagogical concepts for CS education. At the end of each instructional day teachers synthesized how what they learned would help in their classroom. Figure 1 shows an example of the assignment question. Figure 2 shows an example of the collected teachers' assignment writings.

3.5. Data demographic

We collected teachers' assignment writings for every teacher. The assignment writings had consistent objectives across all cohorts that followed a common syllabus despite the


Comparing CS Content in the Classroom

Each day, we will be going through a computer science topic from four different perspectives:

- The concept as an idea in and of itself
- An elementary level approach
- A middle school level approach
- A high school level approach

For these topics, you will be tasked with identifying and comparing what the primary focus of the concept is at each level. Additionally, you should reflect on how understanding the range of levels can help influence your own teaching at your grade level.

For these reflections, use the template below.

Figure 1. The assignment question.

Topic: Conditionals

Level	Reflect on differences and similarities between core focus of the topic, language used, activities, complexity, etc.
Elementary	Learning the basic concepts of if/then, if/else. Using concrete examples to represent conditionals (if it's raining outside, then we have indoor recess) Lots of good activities to represent conditionals (snowman building) Blockly does a nice job presenting conditionals in an easy to see way
Middle	Same concepts as elementary, but also introduces else if and booleans Flowcharts as a visual representation of conditional logic Using math inside of conditionals (if x is < y then…)
High	Continues to expand on what is done at the middle school level, but adds truth tables and more complex comparisons. DeMorgan's Laws "Not A and Not B is the same as Not (A or B)"

How will knowledge of this topic at varying levels impact your teaching strategies in your own classroom?

Conditionals are one of the "easier" concepts to introduce to students simply due to the amount of concrete examples that are available to represent what a conditional does. Like I've said previously, knowing what middle school is doing can help prepare my students for what they will see at the next level. Exposing 5th grade to some of the more complex conditionals that use math comparisons to see what they can pull from their math knowledge.

Figure 2. An example of the collected teachers' assignment writings.

specific assignment topic arrangements were slightly different. Table 1 shows the overview of the assignments for each cohort. There are 76 participant teachers across four cohorts. On average, 1,725.36 words are collected per teacher with a standard deviation of 725.47 words.

Participants were 76 teachers across four cohorts (Nugent et al., 2020). These participants can be grouped based on six demographic categories shown in Table 2: (1) school types, (2) teaching grades, (3) teaching subjects, (4) biological gender, (5) teaching experience, and (6) CS teaching experience (Morrow et al., 2021; Nugent et al., 2020). Additionally, we apply ANOVA (analysis of variance) to

Table 1. An overview of the data collected in the two cohorts. *Cohort 1 introduces CT and pedagogical content, while it had no assignment in the cohort configuration.

	Assignments	
	CS Concepts	CT and Pedagogical Concepts
Cohort 1*	(1) Variables (2) Conditionals (3) Loops (4) Functions	
Cohort 2	(1) Variables (2) Conditionals (3) Loops (4) Functions	(5) Ozobots PBL (6) Dash Cue Teaching Strategies (7) Scratch Flowcharts (8) Differentiation Assessment
Cohort 3	(1) Variables (2) Conditionals (3) Loops (4) Functions	(5) Dash Cue Teaching Strategies (6) Micro:bit teaching strategies
Cohort 4	(1) Variables (2) Conditionals (3) Loops (4) Functions	(5) Dash Cue Teaching Strategies (6) Micro:bit teaching strategies

Table 2. Demographic information about the 76 teachers from the four summer institutes. *Some teachers did not provide their categorical information about teaching grades and thus the sum for the teaching subject categories does not equal the total number of teachers (76).

Categories		Number of Teachers
Summer Institutes	Cohort 1	23
	Cohort 2	23
	Cohort 3	13
	Cohort 4	17
School Types	City/Suburban	33
	Town/Rural	43
Teaching Grades*	Elementary	47
	Middle School	28
Teaching Subjects	Non-STEM	49
	STEM	27
Biological Gender	Female	61
	Male	15
Teaching Experience*	<10 years	20
	≥10 and <20 years	30
	≥20 years	23
CS Teaching Experience*	<5 years	37
	≥5 and <10 years	8
	≥10 years	8

Table 3. Tests of between-subjects effects between school type category and teaching grade category.

Dependent Variable: incremental performance

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	1000.538a	3	333.513	2.285	0.086
Intercept	14253.114	1	14253.114	97.648	<.001
city/suburban vs. town/rural* elementary vs. middle school	910.238	1	910.238	6.236	0.015
city/suburban vs. town/rural	138.221	1	138.221	0.947	0.334
elementary vs. middle school	298.533	1	298.533	2.045	0.157
Error	10509.403	72	145.964		
Total	32549.428	76			
Corrected Total	11509.940	75			

a. R Squared = .087 (Adjusted R Squared = .049)

examine differences in the three categories of teachers' post-PD performance and incremental performance. The ANOVA results (Table 3 and Figure 3) revealed a statistically significant difference between the school type category and the teaching grade category in terms of teachers' incremental performance. This observation indicates that different categories of teachers have different performances during PD, prompting our investigation of whether sentiment and emotion factors affect teachers' performance differently in distinct groups.

4. Investigation

In the following section, we first conduct initial analyses to comprehend the characteristics of the collected data. Second, we report our investigations on the relationships between the sentiment score and teachers' performance (Section 4.2), and those between the emotion score and teachers' performance (Section 4.3), using backward linear

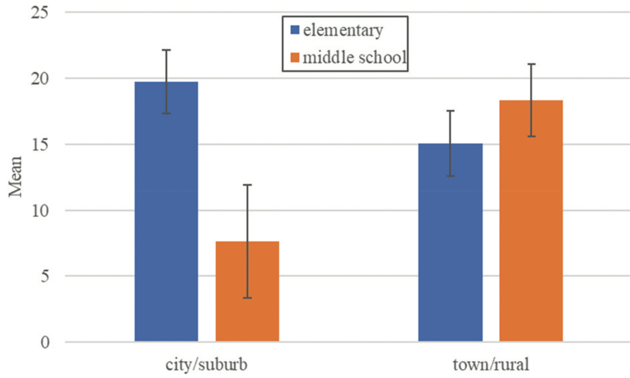


Figure 3. Interaction of school type category (city/suburban vs. town/rural) and teaching grade category (elementary vs. middle school) by two-way ANOVA for the incremental performance of teachers (p = 0.016).

regression. Finally, we report our investigations to discover additional perspectives provided by the sentiment/emotion analysis (Section 4.4) by applying the same backward linear regression between the post-PD survey outcome (Nugent et al., 2020) and teachers' performance, then, comparing it with the findings in Sections 4.2 and 4.3.

4.1. Initial data analyses

Initial data analyses are conducted to comprehend the characteristics of the collected data. The data includes the emotion and sentiment scores (Section 3.1) of teachers along with the information about their demographic groups (Section 3.5).

First, we compute the mean and standard deviation of teachers' emotion and sentiment scores to have a sense of the basic characteristics of these factors, as shown in Table 4. For example, teachers on average had a higher positive sentiment score and a lower negative sentiment score. The highest average emotion is trust while the lowest one is disgust.

Second, we conduct interclass correlation analysis among the emotion and sentiment scores to explore the variable independence. Table 5 shows the interclass correlation coefficient and their significance. The findings indicate that there

Table 4. The mean and standard deviation of sentiment and emotion scores of the 76 participant teachers across the four summer institutes.

	Mean	Standard Deviation
positive sentiment	0.364	0.034
negative sentiment	0.056	0.022
anticipation	0.115	0.030
joy	0.104	0.025
surprise	0.049	0.016
trust	0.217	0.038
disgust	0.008	0.007
fear	0.040	0.015
anger	0.020	0.012
sadness	0.027	0.014

are significant associations between some of the emotion and sentiment scores, implying that these variables are not independent. For example, emotions of joy and trust had a positive correlation of 0.24 with a p-value of 0.038, which suggests a statistically significant positive relationship between these two scores. Also, the negative sentiment score had a statistically significant negative correlation with several other scores (e.g. anticipation, joy, and trust). This analysis supports our use of the backward linear regression analyses to investigate the association between teacher performance and their sentiments and emotions. The interdependencies among sentiment and emotion scores are significant. Utilizing backward linear regression analyses enables us to identify the key scores that account for the variance of teacher performance in sentiment and emotion scores.

Third, we conduct the Shapiro-Wilk test (Shapiro & Wilk, 1965) to explore the variables' normality. The Shapiro-Wilk test is used because our sample size is relatively small (i.e. 76 participant teachers). Table 6 shows the normality test results of the emotion and sentiment scores of the teachers. We found that anticipation, disgust, fear, and anger had normality. Other scores such as joy, surprise, trust, sadness, positive sentiment, and negative sentiment did not show normality.

Fourth, we conduct the Levene test (Levene, 1960) based on mean emotions and sentiments to explore homogeneity of variance across different demographic categories for the emotion and sentiment scores. Table 7 shows the test results of variance homogeneity across six demographic categories including school types, teaching subjects, teaching grades, biological gender, teaching experience, and CS teaching experience. Statistically significant Levene tests ($p < 0.05$) suggest that there are significant differences in variance among the demographic categories for the corresponding emotion and sentiment scores. In the test results, for most emotion and sentiment scores the homogeneity assumption of the variance was met across different demographic categories. Only two scores did not meet the homogeneity assumption of the variance; specifically, the surprise score across the teaching subjects category and the fear score across the teaching grades category vary significantly. These test results are encouraging, as they suggest the viability of using automated sentiment and emotion analysis to study demographic differences in teacher's emotions and sentiments during a PD.

4.2. Investigating the association between sentiments and performance

Our first investigation was designed to discover whether there is a relationship between sentiment scores – and teachers' performance. We carried out seven backward linear regression analyses to investigate associations between sentiment scores (i.e. overall, variability in and trend of sentiment described in Section 3.1) and performance scores (Section 3.2). The seven backward linear regressions are for (1) all 76 teachers in the four cohorts (All), (2) the city/suburban teachers and (3) the town/rural teachers where teachers are grouped by their school types, (4) the elementary teachers and (5) middle school teachers where teachers are grouped by their teaching grades, (6) the non-STEM teachers and (7) STEM teacher where teachers are grouped by their teaching subjects.

Table 5. Interclass correlation analysis among emotion and sentiment affect items using scores generated by NRCLex. *Significant ($p < 0.05$) correlation.

	anticipation	joy	surprise	trust	disgust	fear	anger	sadness	positive sentiment	negative sentiment
anticipation	1.00									
joy	0.28* $p = 0.014$	1.00								
surprise	0.03 $p = 0.784$	0.24* $p = 0.038$	1.00							
trust	−0.35* $p = 0.002$	−0.20 $p = 0.080$	−0.03 $p = 0.808$	1.00						
disgust	−0.15 $p = 0.200$	−0.19 $p = 0.106$	0.03 $p = 0.801$	−0.20 $p = 0.080$	1.00					
fear	−0.13 $p = 0.257$	−0.40* $p = 0.000$	−0.15 $p = 0.205$	−0.41* $p = 0.000$	0.26* $p = 0.023$	1.00				
anger	−0.21 $p = 0.068$	−0.18 $p = 0.122$	−0.18 $p = 0.125$	−0.45* $p = 0.000$	0.41* $p = 0.000$	0.48* p=0.000	1.00			
sadness	−0.31* $p = 0.006$	−0.22 $p = 0.051$	−0.15 $p = 0.195$	−0.35* $p = 0.002$	0.32* $p = 0.005$	0.43* p=0.000	0.46* p=0.000	1.00		
positive sentiment	−0.24* $p = 0.038$	−0.19 $p = 0.108$	−0.35* $p = 0.002$	0.20 $p = 0.081$	−0.46* $p = 0.000$	−0.29* p=0.012	−0.38* p=0.001	−0.34* p=0.003	1.00	
negative sentiment	−0.29* $p = 0.012$	−0.50* $p = 0.000$	−0.19 $p = 0.110$	−0.45* $p = 0.000$	0.50* $p = 0.000$	0.55* p=0.000	0.65* p=0.000	0.60* p=0.000	−0.29* p=0.013	1.00

Table 6. Normality test of each emotion and sentiment score using the Shapiro-Wilk test. *Significant ($p < 0.05$) correlation.

	Shapiro-Wilk	p-value
anticipation	0.96*	0.017
joy	0.99	0.606
surprise	0.98	0.360
trust	0.98	0.162
disgust	0.87*	0.000
fear	0.97*	0.043
anger	0.96*	0.025
sadness	0.97	0.092
positive sentiment	0.99	0.941
negative sentiment	0.99	0.952

4.2.1. Sentiment analysis for all teachers

Backward linear regression did not identify any statistically significant associations between teachers' sentiments – in terms of overall, variability in, and trend of – and their performance scores for all teachers in the four cohorts. The sentiment has two polarities: positive and negative that represent the overall attitude of teachers toward the topic. No statistically significant association is identified. We suspect that this is because the overall attitude is too general for the all-teacher sentiment analysis.

4.2.2. Sentiment analysis for teachers of different demographic groups

Table 8 shows the identified significant associations between teachers' sentiments and their performance scores for all teachers grouped by three pairs of categories (School Types, Teaching Grades, Teaching Subjects in Table 2) using backward linear regression. In the analysis for all teachers (Section 4.2.1), we hypothesized that the overall attitude could be too general for all-teacher analysis to show meaningful results. Indeed, by looking at the teachers by different demographic groups, we found significant associations between sentiment scores and teachers' performance for different groups of teachers. For example, backward linear regression identified significant associations between the overall positive sentiment and post-PD performance for non-STEM teachers. Thus, we are encouraged that the sentiment analysis based on teacher-submitted assignments appears to have relationships with their performance during PD.

Further, we see that both variability in and trend of teachers' sentiments can provide additional nuances to the overall sentiment to predict teachers' performance to help them complete their PD more successfully. Also as shown in Table 8, backward linear regression identified significant associations between the variability in sentiment and teachers' performance, and between the trend of sentiment and teachers' performance provided additional insights for city/suburban, town/rural, non-STEM, elementary, and middle school teachers.

In summary, we see that one could monitor and look for such sentiment scores in a teacher's submitted assignments on successive days to prompt intervention or inquiries with the teacher to address potential concerns for the teacher. For example, if we observed a high variability in negative sentiment or a high trend of positive sentiment for elementary teachers in their written assignments, then, these elementary teachers are

Table 7. Variance homogeneity test of emotion and sentiment scores across different demographic categories using the Levene test. *Significant ($p < 0.05$) correlation.

	School Types		Teaching Subjects		Teaching Grades		Biological Gender		Teaching Experience		CS Teaching Experience	
	Levene	p-value	Levene	p-value	Levene	p-value	Levene	p-value	Levene	p-value	Levene	p-value
anticipation	0.30	0.583	0.01	0.919	0.92	0.341	0.04	0.836	1.01	0.371	0.69	0.508
joy	0.06	0.806	0.77	0.382	0.06	0.800	1.02	0.316	1.94	0.151	0.15	0.859
surprise	0.02	0.902	4.88*	0.030	1.38	0.244	0.63	0.431	0.53	0.589	0.70	0.499
trust	1.35	0.248	0.07	0.794	1.22	0.274	1.64	0.204	0.11	0.893	1.09	0.344
disgust	0.71	0.402	2.99	0.088	0.82	0.368	1.47	0.229	0.93	0.398	0.05	0.954
fear	2.52	0.117	1.20	0.277	5.86*	0.018	0.81	0.372	0.23	0.796	1.24	0.297
anger	2.84	0.096	0.00	0.978	0.96	0.329	0.00	0.994	0.11	0.895	1.62	0.208
sadness	0.26	0.612	0.03	0.863	3.69	0.059	2.31	0.133	0.51	0.605	0.42	0.662
positive sentiment	1.45	0.233	3.70	0.058	1.26	0.265	0.01	0.911	2.53	0.087	0.94	0.399
negative sentiment	0.00	0.972	1.74	0.191	0.11	0.740	0.05	0.823	0.30	0.742	0.44	0.647

Table 8. The backward linear regression analyses between teachers' sentiments and their performance scores for 76 teachers grouped by the demographic category. *Significant (p < 0.05).

	Post-PD Performance		Incremental Performance		Post-PD Performance		Incremental Performance	
	City/Suburban (N=33)				**Town/Rural (N=43)**			
Overall Variability in Trend of	negative sentiment	505.44 $p = 0.013$			negative sentiment	−149.81 $p = 0.050$		
	Non-STEM ($N = 47$)				**STEM ($N = 28$)**			
Overall Variability in Trend of	positive sentiment	132.89 $p = 0.017$						
	negative sentiment	−134.54 $p = 0.041$						
	Elementary ($N = 49$)				**Middle School ($N = 27$)**			
Overall Variability in	negative sentiment	−179.19 $p = 0.009$	negative sentiment	−190.06 $p = 0.008$			positive sentiment	119.36 $p = 0.029$
Trend of			positive sentiment	−159.21 $p = 0.049$				

Table 9. Mapping between emotion affect items in NRCLex emotion analysis and the psychological constructs presented as part of the achievement emotions questionnaire.

NRCLex Emotion Affect Items	Achievement Emotions Questionnaire (AEQ)
anticipation	Hope
joy	Enjoyment
surprise	
trust	
	Pride
disgust	Shame
fear	Anxiety
anger	Anger
sadness	Hopelessness
	Boredom

likely having challenges, changes, or frustrations affecting their performance. Subsequently, one can intervene and provide additional support.

4.3. Investigating the association between emotions and performance

In this investigation, we report our association analyses between emotions (i.e. overall emotions and standard deviation of emotions described in Section 3.1) and performance scores (Section 3.2), similar to those reported in Section 4.2.

We further map the eight emotion items from the NRCLex emotion analysis to the psychological constructs Pekrun and his colleagues (2011) presented as part of the Achievement Emotions Questionnaire (AEQ) to help us interpret the results. As shown in Table 9, while the mapping between the emotion items and the constructs is not complete, there are certainly overlaps between the two sets. Note that in AEQ, hope and enjoyment are positive activating emotions, anger, anxiety, and shame are negative activating emotions, and hopelessness is a negative deactivating emotion.

We also perform an internal consistency analysis on these eight emotion affect items. Specifically, we identify the relationship between each pair of emotion affect items based on the statistically significant correlations (i.e. p-value < 0.05) that appeared in the analysis. As shown in Table 5, we see that (1) the emotions anticipation (representing Hope of AEQ) and joy (representing Enjoyment of AEQ) are positively correlated, and (2) the emotions anger (representing Anger of AEQ), fear (representing Anxiety of AEQ), and disgust (representing Shame of AEQ) are positively correlated to each other. This shows that the NRCLex-generated emotion affect scores are consistent.

Table 10. The backward linear regression analyses between teachers' emotions and performance for the 76 teachers from the four cohorts. *Significant ($p < 0.05$).

	Post-PD Performance		Incremental Performance	
Overall			trust	−83.08 $p = 0.040$
			anger	−318.74 $p = 0.010$
Variability in	trust	−83.62 $p = 0.011$		
Trend of			anticipation	−218.59 $p = 0.023$

4.3.1. Emotion analysis

Table 10 shows the identified significant associations between teachers' emotion scores and their performance scores using backward linear regression. The emotion analysis using the eight emotion affect items is a finer grained analysis since the positive sentiment and the negative sentiment scores are compounds of positive emotions and negative emotions. Backward linear regression identified statistically significant associations between the variability in emotion – trust—score and teachers' post-PD performance, between the overall emotion – trust and anger – scores and teachers' incremental performance, and between the trend of emotion – anticipation—score and teachers' incremental performance for all teachers. This encourages us that the emotion analysis could also provide finer insights than the sentiment analysis for different demographic groups of teachers. Further, these identified associations indicated that the emotion analysis based on teacher-submitted assignments appears to be related to teachers' performance. The overall, variability in, and trend of emotions can be used to identify the challenges, changes, and frustrations that affect teachers' performance. Thus, similar to our findings from the sentiment analysis, the emotion analysis could also help us administer intervention or carry out inquiries with the teacher to address potential concerns.

Further, to investigate whether the findings of variability in emotions were affected by the average intensity of the emotion scores involved in the computation of the variability, we also carry out additional analysis for the variability in trust above that has been identified as significantly associated with teachers' performance. Specifically, the analysis first groups the 76 teachers into two subsets based on the average intensity of the emotion scores involved in the computation of the variability: (1) lower average, and (2) higher average. Specifically, in the lower average group, teachers have their emotion score intensities smaller than the average emotion score intensities of all 76 teachers. Similarly, in the higher average group, teachers have their emotion score intensities larger than the average emotion score intensities of all 76 teachers. Then, we perform the correlation analysis between the variability in the emotion and teachers' performance to observe if the significance of the correlation changed compared to that found in the all-teacher analysis. Table 11 shows that the significance of the correlation between the variability in trust and teachers' performance diminishes for teachers with a lower average intensity of the emotion scores. Meanwhile, the significance of the correlation coefficient remained for the variability in trust for teachers with a higher average intensity of the emotion scores. These observations suggest that variability in emotions is likely more effective to monitor for teachers who have a high average intensity of the emotion scores.

Table 11. The correlation analysis between variability in teachers' trust emotion and their performance for the 76 teachers from the four cohorts grouped by the average of variability and mean intensity of the emotion. *Significant ($p < 0.05$) correlation.

			Correlation between variability in emotion and performance	
			Post-PD Performance	Incremental Performance
Average intensity of the emotion scores	Lower	trust (N=41)	−0.20 $p = 0.21$	−0.05 $p = 0.77$
	Higher	trust (N=35)	−0.33* $p = 0.05$	−0.23 $p = 0.18$

Table 12. The backward linear regression analyses between teachers' emotions and their performance for 76 teachers grouped by demographic categories. *Significant (p < 0.05).

	Post-PD Performance		Incremental Performance		Post-PD Performance		Incremental Performance	
	City/Suburban (N=34)				**Town/Rural (N=42)**			
Overall			trust	-200.91 $p = 0.000$	disgust	-784.55 p=0.043	anticipation	195.33 p=0.010
			fear	-789.34 $p = 0.000$				
Variability in			anticipation	-238.19 $p = 0.009$	joy	-217.21 p=0.041	anticipation	-185.11 $p = 0.041$
					disgust	537.92 p=0.027		
Trend of			joy	-377.23 $p = 0.007$				
			trust	-236.44 $p = 0.005$				
	Non-STEM ($N = 48$)				**STEM (N=28)**			
Overall	joy	-189.20 $p = 0.001$			joy	305.06 $p =$ 0.009		
	sadness	-379.04 $p = 0.001$			sadness	392.15 $p = 0.042$		
Variability in	anger	-161.54 $p = 0.024$			anticipation	272.46 $p = 0.033$		
					trust	-343.13 $p = 0.015$		
Trend of					anticipation	-454.10 $p = 0.038$		
	Elementary ($N = 50$)				**Middle School ($N = 26$)**			
Overall	anger	-289.19 $p = 0.016$	anticipation	101.12 $p = 0.050$			surprise	515.87 $p = 0.005$
							trust	-131.26 $p = 0.031$
							disgust	$-2,760.87$ $p = 0.001$
Variability in	trust	-105.31 $p = 0.001$	trust	-79.80 $p =$ 0.006			anticipation	326.87 $p = 0.024$
							joy	-803.39 $p = 0.002$
							surprise	-653.00 $p =$ 0.000
							trust	292.00 $p = 0.012$
							disgust	1,562.63 $p = 0.012$
							fear	398.46 p=0.020
							anger	$-1,035.97$ $p = 0.001$
Trend of	anger	-418.78 $p = 0.034$	anticipation	-328.43 $p = 0.006$			anticipation	726.88 $p = 0.007$
	sadness	426.05 $p =$ 0.016					joy	-754.12 $p = 0.024$
							surprise	589.65 $p = 0.012$
							disgust	$-4,556.62$ $p = 0.008$
							fear	$-1,726.39$ $p = 0.001$
							anger	2,713.41 $p = 0.001$

4.3.2. Emotion analysis for teachers of different demographic groups

Table 12 shows the identified significant associations between teachers' emotions and their performance scores for all teachers grouped by three pairs of categories (School Types, Teaching Grades, Teaching Subjects in Table 2) using backward linear regression. We observe statistically significant associations between multiple overall emotion scores and teachers' performance across the three pairs of categories. The directions of these associations are consistent with the AEQ emotion mapping in Table 9 except for non-STEM and STEM teachers. Specifically, anticipation mapping to hope of AEQ are positive activating emotions, which are tied to strengthening motivational processes (Pekrun et al., 2011). Anger, fear, and disgust mapping to anger, anxiety, and shame of AEQ are negative activating emotions, which can undermine learning motivation (Pekrun et al., 2011). The above observations indicate the viability of the emotion analysis, that it is able to "measure" emotions – using teachers' assignments – that are consistent with their performance.

In addition, we found for non-STEM and STEM teachers, joy and sadness indicated different affects on teachers' performance. Specifically, joy and sadness indicated associations with improvement in teachers' post-PD performance for STEM teachers while they indicated associations with a decrement in teachers' post-PD performance for non-STEM teachers. We suspect that this observation is rooted in the difference between non-STEM and STEM teachers. Non-STEM and STEM teachers could have different mindsets of learning knowledge during PD.

Further, we also found that the variability in and trend of emotions provide additional nuances to the overall emotion, which is similar to the effect found of variability in sentiment and trend of sentiment to the overall sentiment (Section 4.2.2). Specifically, 16 additional significant associations between the variability in emotion and teachers' performance, and 12 additional significant associations between the trend of emotion and teachers' performance were shown in Table 12. Especially for middle school teachers, 7 out of 8 variability in emotion items and 6 out of 8 trend in emotion items were presented in the result of backward linear regression, which provided a very strong model indicating associations between emotions and the performance of teachers.

Moreover, combining the sentiment analysis (Table 8) and the emotion analysis (Table 12) can provide us with a finer set of insights to identify the challenges, changes, and frustrations of teachers in order to help them perform well during PD. We observed that at least three sentiment or emotion affect items can be used as indicators for all investigated teacher groups. Thus, one can leverage these sentiment/emotion indicators to better understand and model teachers' performance in a PD for all of the three categories of teacher groups.

4.4. Additional perspectives provided by the sentiment and emotion analyses

In this investigation, we aimed to uncover additional perspectives provided by the sentiment and emotion analyses. We compared the associations found for sentiment and emotion analysis with the association between six post-PD survey outcomes (Nugent et al., 2020) and performance scores. The survey outcomes consist of six numerical scores:

Table 13. The backward linear regression analyses between teachers' post-PD survey outcomes and their performance for the 76 teachers from the four cohorts. *Significant (p < 0.05).

Post-PD Performance		Incremental Performance
CS Content	0.21 $p = 0.009$	

Table 14. The backward linear regression analyses between teachers' post-PD survey outcomes and their performance scores for 76 teachers grouped by the demographic category. *Significant (p < 0.05).

City/Suburban (*N*=33)			**Town/Rural (*N*=43)**		
Post-PD Performance		Incremental Performance	Post-PD Performance		Incremental Performance
Non-STEM (*N* = 47)			**STEM (*N* = 28)**		
Personal interest	8.91 $p = 0.003$				
Elementary (*N* = 49)			**Middle School (*N* = 27)**		
CS Content	0.27 $p = 0.006$		Perceived Value	21.06 $p = 0.037$	
Personal interest	5.68 $p = 0.044$				

(1) CS self-efficacy, (2) CS content, (3) CS pedagogy, (4) CS attitudes, (5) personal interest, and (6) perceived value. CS self-efficacy was determined through a project-developed 31-item confidence instrument measuring two constructs: (1) CS pedagogy (16 items) and (2) CS content (6 items). Items were rated on a 0–100% confidence scale and were developed to align with objectives of each of the summer courses. CS attitudes, personal interest, and perceived value were measured using a Likert scale (1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree). The teacher instrument was developed by adapting the Computing Attitudes Survey (Dorn & Elliott Tew, 2015), which was validated with CS undergraduates. More details about the survey outcomes can be found in the work of Nugent et al. (2020). We report this investigation for both all-teacher analysis and grouped teacher analysis similar to those reported in Sections 4.2 and 4.3 using backward linear regression.

4.4.1. All-teacher post-PD survey analysis

Table 13 shows the observed significant association between teachers' post-PD survey outcomes and their performance for all teachers. Backward linear regression identified the CS Content outcome was significantly associated with teachers' post-PD performance. There were the overall and the trend of emotions statistically significantly associated with teachers' incremental performance (Section 4.3.1), which hinted that emotions provided additional perspectives indicating teachers' performance and played an important role during PD. We see that the associations of emotions can supplement survey-based analysis.

4.4.2. Different demographics post-PD survey analysis

Table 14 shows the identified significant associations between teachers' post-PD survey outcomes and their performance scores for all teachers grouped by three pairs of categories (School Types, Teaching Grades, Teaching Subjects in Table 2) using backward linear regression. In comparison with the association found in sentiment and emotion analyses, we identified two additional perspectives as the following.

(1) Additional sentiment/emotion scores appeared to be associated with teachers' incremental performance for city/suburban teachers, town/rural teachers, elementary teachers, and middle school teachers while no significant associations were observed in the post-PD survey analysis.
(2) Additional sentiment/emotions scores were found to be associated with teachers' post-PD performance for city/suburban teachers, town/rural teachers, and STEM teachers while no significant associations were observed in the post-PD survey analysis for these teachers.

Thus, the sentiment and emotion analyses based on teacher-submitted assignments appear to provide additional perspectives to the conventional post-PD survey during PD. That is, one could utilize the proposed sentiment and emotion analyses as a supplemental analysis to follow up with teachers during PD.

5. Limitations of findings

This investigation has three limitations. First, the sentiment and emotion analysis tool, NRCLex, may not be highly accurate for the PD assignment writings. On one hand, NRCLex is based on a lexicon dictionary that is built using crowdsourcing (Mohammad & Turney, 2013). The sentiment and emotion scores of some words within NRCLex may raise questions. For example, the word "cable" is labelled as associated with the surprise emotion, while the definition of "cable" shows as a neutral word in the WordNet lexicon database (Princeton University, n.d..). A further investigation and possibly correction to fine-tune the dictionary based on teachers' writings on the assignment could benefit to obtain more accurate sentiment and emotion scores. On the other hand, teachers' writings to the PD assignments could benefit from longer text lengths of their assignment writings. The sentiment and emotion scores are estimated based on the word-sentiment association dictionary and the word-emotion association dictionary of NRCLex. For example, if words associated to the trust emotion appeared in texts, then the texts are associated with the trust emotion as well. The estimation considers all sentiment and emotion associations for each word, then computes a joint score for the texts. Further, if words associated with an emotion affect item (e.g. the anger emotion) were missing from the texts, then the texts are not associated with the missing emotion affect items. Thus, the longer the texts, the more the possibility that the texts can contain words for all emotion affect items for fuller sentiment and emotion analyses. Nevertheless, in our investigation, we observed 23 out of 760 (~3%) sentiment and emotion scores returned zero on some items for teachers. We consider that the number of words in teachers' assignment writings is adequate for our investigation since at least 7 out of 10 emotion scores have been successfully estimated (i.e. non-zero scores) for all teachers. However, we see that it would be beneficial to collect more writings from teachers to have all estimated scores for all sentiment and emotion items for a fuller analysis.

Second, there are opportunities to further enrich the data used in our investigations. Presently, in our investigation, we used data collected from the four years of two-week summer institutes during PD. We did not use, however, data from other elements of the PD such as the five workshops conducted during the academic year for each cohort as well as the online discussions involved among the teachers. These data reflect how teachers apply the content of PD to their classrooms. We could use sentiment and emotion analyses to estimate scores for these data such as feedback and comments from their classroom teaching to investigate the relationship between teachers' practicing in classrooms and their PD performance evaluation. Then, the investigation would provide additional perspective between teachers practice what they have learned in their classrooms and their PD performance during the summer courses. These additional perspectives could further strengthen our findings of sentiment and emotion analyses for PD.

There are potentials for further exploration into the influence of instructors on PDs based on relationships between teachers' performance and their sentiments and emotions. In our investigation, we conducted a cohort-wise analysis to examine variations among teachers in different cohorts. We did not observe statistically significant differences in teachers' performance and their sentiment and emotion scores when comparing each pair of cohorts using the student t-test (all p-values > 0.05). However, we did observe statistically significant differences in the correlation coefficient between teachers' performance and their sentiment and emotion scores between Cohort 1 and the other cohorts (with a p-value < 0.05 when comparing Cohort 1 to each of Cohorts 2–4). This observation may be attributed to Cohort 1 having a different instructor from that for the other three cohorts, despite all cohorts following a common syllabus. This suggests a further investigation into the relationship between teachers' performance and their sentiment and emotion scores could yield additional insights on role of instructors for PDs.

6. Conclusion and future work

In this investigation, we have investigated the relationship between teachers' sentiment and emotion scores, and their performance to provide additional analysing perspectives for conventional survey approaches in professional development (PD). Specifically, the investigation estimated sentiment and emotion scores based on teachers' writing to an assignment using a toolkit, namely NRCLex (Mohammad & Turney, 2013). Then, the investigation evaluated associations between the sentiment and emotion scores and teachers' PD performance using backward linear regression. In conclusion, we have found that sentiment and emotion analyses based on teachers' writings on an assignment can be used to help gauge teachers' learning performance during PD to complement conventional PD surveys. Further, in general, we identify two suggestions to utilize sentiment and emotion analysis to gauge teachers during PD:

(1) *One can leverage different sentiment/emotion factors to better understand and model teachers' performance in a PD for different groups of teachers*. For example, as shown in Table 12, we observed that the higher the joy emotion score the higher the post-PD performance for STEM teachers, while the lower the joy emotion score the higher the post-PD performance for non-STEM teachers. One can, based on these

observations, design an intervention to reduce the joy emotion in non-STEM teachers and increase the joy emotion in STEM teachers in order to help improve teachers' final PD performance.

(2) *Sentiment and emotion analysis can supplement conventional post-PD survey outcomes by providing a different perspective correlated to teacher performance*. For example, in group-based analysis (Sections 4.2.2 and 4.3.2), there are statistically significant associations between sentiment/emotion scores and teachers' performance, while there is no statistically significant association between the conventional survey outcome and teachers' performance for city/suburban teachers, town/rural teachers, and STEM teachers.

Further, we draw two insights based on this investigation:

(1) *More timely intervention strategies can be deployed based on sentiment and emotion analyses for PD*. Based on the daily assignment submitted by teachers, the sentiment and emotion analyses can be used to monitor and provide a daily estimation or "prediction" of the challenges, changes, or frustrations faced by teachers affecting their performance. Thus, we could discover teachers' issues in a more timely manner and intervene to improve the effectiveness of PD.
(2) *Variability in and trend of sentiments and emotions can provide additional insights that a conventional pre- and post-PD survey approach cannot provide*. Our investigation has shown that variability in and trend of certain sentiment and emotion factors is associated with teachers' performance. They represent the teachers' affective changes in their emotions and sentiments during PD, while a conventional PD survey only reflects teachers' status at a single time point (e.g. pre-PD or post-PD). We see that the sentiment and emotion analyses based on teachers' writings could be used to enrich our understanding and modelling of teachers' performance.

We identify four recommendations for organizing a similar PD for CS education based on our investigation. First, PD organizers can gain a day-to-day understanding of teachers' sentiments and emotions. This knowledge can inform timely adjustments to PD content and delivery. For example, if teachers showed low joy emotion, organizers could add activities to increase teacher enjoyment with the PD. Second, organizers can identify teachers who might be struggling and provide timely emotional support. For example, if an individual is feeling overwhelmed (e.g. high negative sentiment), they can arrange one-on-one meetings or offer additional attention to help that individual engage more effectively in discussions. Third, organizers can form groups of teachers to facilitate peer support. For example, teachers with more experience can be paired with those expressing frustration to share insights and provide encouragement within the cohort. Fourth, organizers can adjust their communication and support strategies accordingly based on sentiments and emotions. The awareness of timely sentiments and emotions can lead to more effective interactions with teachers and an overall improved PD experience. For example, if organizers notice that a group of teachers has shown sentiment or emotion scores indicating performance improvement enthusiasm on a topic during PD, they can adjust their communication to encourage further discussion and collaboration on that topic.

Next, we plan to expand our investigation into five aspects. First, because more data samples could benefit to strengthen our findings as shown in Section 5, we plan to include more data – such as the data from the workshops and discussions during the academic year of the PD – to further strengthen the findings our investigations or identify nuances with the different emotion and sentiment attributes. Second, we plan to investigate what and how to intervene when we detect, say, variability in teachers' emotions and sentiments to improve their performance during a PD, as an application of the findings reported in this paper. Third, we plan to investigate how to use sentiment and emotion analyses to supplement the conventional survey to help improve teachers' performance during PD and the effectiveness of PD. Fourth, we plan to investigate additional factors impacting teachers' sentiments and emotions during PD, such as their instructional styles (e.g. the three approaches reported by (Searle et al., 2023): direct instruction, discovery learning, and scaffolding and modelling), and views of algorithms (e.g. the two views reported by (Nijenhuis-Voogt et al., 2021, 2023): focused on "thinking" or focused on "thinking and making"). Finally, we plan to further expand our investigation to other PDs or classrooms. The text-based emotion and sentiment analysis method we have employed in our investigation is not limited to CS teachers within PDs. For example, instructors in secondary and post-secondary classes could employ the method to gain timely insights into students' emotions and sentiments during their classes, allowing them to promptly identify any challenges students may be facing and implement effective remedies.

Acknowledgments

This work was supported by a National Science Foundation grant no. 1837476. All findings and opinions are those of the authors and not necessarily of the funding agency. The authors would like to acknowledge the instructors and teaching assistants and teachers of the four summer institutes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Science Foundation [1837476].

ORCID

Yi Liu http://orcid.org/0000-0002-0879-0487

References

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An easy-to-use framework for state-of-the-art NLP. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, 54–59. https://doi.org/10.18653/v1/N19-4010

Allison, J. (2023). Classifying the characteristics of effective Continuing Professional Development (CPD) for computer science teachers in the 16-18 sector. *ACM Transactions on Computing Education*, *23*(2), 1–30. https://doi.org/10.1145/3582275

Anastasiadou, S. D., & Karakos, A. S. (2011). The beliefs of electrical and computer engineering students' regarding computer programming. *The International Journal of Technology, Knowledge, and Society*, *7*(1), 37–51. Scopus. https://doi.org/10.18848/1832-3669/cgp/v07i01/56170

Bandura, A., & Watts, R. E. (1996). Self-efficacy in changing societies. *Journal of Cognitive Psychotherapy*, *10*(4), 313–315. https://doi.org/10.1891/0889-8391.10.4.313

Biswas, G., Hutchins, N., Lédeczi, Á., Grover, S., & Basu, S. (2019). Integrating computational modeling in K-12 STEM classrooms. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 1288. https://doi.org/10.1145/3287324.3293757

Bosch, N., & D'Mello, S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education*, *27*(1), 181–206. https://doi.org/10.1007/s40593-015-0069-5

Broneak, C., & Rosato, J. (2021). Developing empathy and persistence through professional development in new to CSA teachers. *Proceedings of the 17th ACM Conference on International Computing Education Research*, 443. https://doi.org/10.1145/3446871.3469793

Cabada, R. Z., Estrada, M. L. B., Hernández, F. G., Bustillos, R. O., & Reyes-García, C. A. (2018). An affective and web 3.0-based learning environment for a programming language. *Telematics and Informatics*, *35*(3), 611–628. https://doi.org/10.1016/j.tele.2017.03.005

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Damasio, A. R. (2004). Emotions and feelings: A neurobiological perspective. In A. Fischer, A. S. R. Manstead, & N. Frijda (Eds.), *Feelings and emotions: The Amsterdam symposium* (pp. 49–57). Cambridge University Press. https://doi.org/10.1017/CBO9780511806582.004

Dawson, T. (2021). The block-based, Text-based, and the CS1 prepared. *Proceedings of the 17th ACM Conference on International Computing Education Research*, 413–414. https://doi.org/10.1145/3446871.3469777

Dolianiti, F. S., Iakovakis, D., Dias, S. B., Hadjileontiadou, S., Diniz, J. A., & Hadjileontiadis, L. (2019). Sentiment analysis techniques and applications in education: A survey. In M. Tsitouridou, J. A. Diniz, & T. A. Mikropoulos (Eds.), *Technology and innovation in learning, teaching and education* (pp. 412–427). Springer International Publishing. https://doi.org/10.1007/978-3-030-20954-4_31

Dorn, B., & Elliott Tew, A. (2015). Empirical validation and application of the computing attitudes survey. *Computer Science Education*, *25*(1), 1–36. https://doi.org/10.1080/08993408.2015.1014142

Faherty, R., Nolan, K., & Quille, K. (2020). A collaborative online micro: Bit K-12 Teacher PD workshop. *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 307. https://doi.org/10.1145/3372782.3408113

Falkner, K., Vivian, R., & Williams, S.-A. (2018). An ecosystem approach to teacher professional development within computer science. *Computer Science Education*, *28*(4), 303–344. https://doi.org/10.1080/08993408.2018.1522858

Fellous, J. M., & Ledoux, J. E. (2005). Toward basic principles for emotional processing: What the fearful brain tells the robot. In Fellous, J. M., & Arbib, M. A., (Eds.), *Who needs emotions?: The brain meets the robot* (pp. 79–115). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195166194.003.0004

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, *20*(2), 303–315. https://doi.org/10.1086/209351

Grover, S., Fisler, K., Lee, I., & Yadav, A. (2020). Integrating computing and computational thinking into K-12 STEM learning. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 481–482. https://doi.org/10.1145/3328778.3366970

Hamlen Mansour, K., Jackson, D. K., Bievenue, L., Voight, A., & Sridhar, N. (2023). Understanding the impact of peer instruction in CS principles teacher professional development. *ACM Transactions on Computing Education*, *23*(2), 1–24. https://doi.org/10.1145/3585077

Hjorth, A. (2021). NaturalLanguageProcesing4All: - a constructionist NLP tool for scaffolding students' exploration of text. *Proceedings of the 17th ACM Conference on International Computing Education Research*, 347–354. https://doi.org/10.1145/3446871.3469749

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web & Social Media*, *8*(1), 216–225. Article 1 https://doi.org/10.1609/icwsm.v8i1.14550

Informatics Europe. (2016). *Informatics for all*. https://www.informatics-europe.org/

Jocius, R., Joshi, D., Dong, Y., Robinson, R., Cateté, V., Barnes, T., Albert, J., Andrews, A., & Lytle, N. (2020). Code, connect, create: The 3C professional development model to support computational thinking infusion. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 971–977. https://doi.org/10.1145/3328778.3366797

Kafai, Y. B., Baskin, J., Fields, D., Goode, J., Twarek, B., & Yadav, A. (2020). Looking ahead: Professional development needs for experienced CS teachers. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 1118–1119. https://doi.org/10.1145/3328778.3366977

Kaya, E., Newley, A., Yesilyurt, E., & Deniz, H. (2021). Nature of computer science: Identification of K-12 accessible nature of computer science tenets and development of an open-ended nature of computer science instrument. *Proceedings of the 17th ACM Conference on International Computing Education Research*, 426. https://doi.org/10.1145/3446871.3469784

Kinnunen, P., & Simon, B. (2012). My program is ok – am I? Computing freshmen's experiences of doing programming assignments. *Computer Science Education*, *22*(1), 1–28. https://doi.org/10.1080/08993408.2012.655091

Kuo, F.-Y., Wu, W.-H., & Lin, C. S. (2013). An investigation of self-regulatory mechanisms in learning to program visual basic. *Journal of Educational Computing Research*, *49*(2), 225–247. https://doi.org/10.2190/EC.49.2.f

Law, K. M. Y., Lee, V. C. S., & Yu, Y. T. (2010). Learning motivation in e-learning facilitated computer programming courses. *Computers & Education*, *55*(1), 218–228. https://doi.org/10.1016/j.compedu.2010.01.007

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 278–292). Stanford University Press.

Liao, Y. W, Huang, Y. M., Huang, Y. M., Su, Z. Y., & Wei, C. W. (2019). An empirical evaluation of online cooperative programming platforms based on the PACT framework and technology readiness. *Journal of Internet Technology*, *20*(2), 345–352. https://doi.org/10.3966/160792642019032002003

Lishinski, A., Yadav, A., & Enbody, R. (2017). Students' emotional reactions to programming projects in introduction to programming: Measurement approach and influence on learning outcomes. *Proceedings of the 2017 ACM Conference on International Computing Education Research*, 30–38. https://doi.org/10.1145/3105726.3106187

Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., Childs, W., Schnurr, J., Qalieh, A., Ragnarsson, L., Coe, J., Calvo, A. L., Kulshrestha, N., Eslava, J., Jcalbert, Harden, T. J., Pavelmalai Kolb, J., Ong, D., & Rachum, R. (n.d.). *TextBlob: Simplified Text Processing—TextBlob 0.16.0 documentation*. Retrieved June 2, 2023, from https://textblob.readthedocs.io/en/dev/#

Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2009). *Designing professional development for teachers of science and mathematics*. Corwin Press.

Magerko, B., Freeman, J., Mcklin, T., Reilly, M., Livingston, E., Mccoid, S., & Crews-Brown, A. (2016). EarSketch: A STEAM-Based approach for underrepresented populations in high school computer science education. *ACM Transactions on Computing Education*, *16*(4), 1–14. https://doi.org/10.1145/2886418

Malmi, L., Sheard, J., Kinnunen, P., Simon, & Sinclair, J. (2020). Theories and models of emotions, attitudes, and self-efficacy in the context of programming education. *Proceedings of the 2020 ACM*

Conference on International Computing Education Research, 36–47. https://doi.org/10.1145/3372782.3406279
Margolis, J., Ryoo, J., & Goode, J. (2017). Seeing myself through someone Else's eyes: The value of In-classroom coaching for computer science teaching and learning. *ACM Transactions on Computing Education*, *17*(2), 1–6. https://doi.org/10.1145/2967616
Menekse, M. (2015). Computer science teacher professional development in the United States: A review of studies published between 2004 and 2014. *Computer Science Education*, *25*(4), 325–350. https://doi.org/10.1080/08993408.2015.1111645
Mohammad, S. M. (2021). Chapter 11 - sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. In H. L. Meiselman (Ed.), *Emotion measurement* (2nd ed., pp. 323–379). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-821124-3.00011-9
Mohammad, S. M. (2022). Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, *48*(2), 239–278. https://doi.org/10.1162/coli_a_00433
Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34.
Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x
Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
Morrow, P., Soh, L.-K., Nugent, G., Smith, W., Trainin, G., & Steen, K. (2021). SWOT analysis of two different designs of Summer professional development institutes for K-8 CS teachers. *2021 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE49875.2021.9637451
Mouza, C., Sheridan, S., Lavigne, N. C., & Pollock, L. (2023). Preparing undergraduate students to support K-12 computer science teaching through school-university partnerships: Reflections from the field. *Computer Science Education*, *33*(1), 3–28. https://doi.org/10.1080/08993408.2021.1970435
Ni, L., Bausch, G., & Benjamin, R. (2023). Computer science teacher professional development and professional learning communities: A review of the research literature. *Computer Science Education*, *33*(1), 29–60. https://doi.org/10.1080/08993408.2021.1993666
Nijenhuis-Voogt, J., Bayram-Jacobs, D., Meijer, P. C., & Barendsen, E. (2021). Omnipresent yet elusive: Teachers' views on contexts for teaching algorithms in secondary education. *Computer Science Education*, *31*(1), 30–59. https://doi.org/10.1080/08993408.2020.1783149
Nijenhuis-Voogt, J., Bayram-Jacobs, D., Meijer, P. C., & Barendsen, E. (2023). Teaching algorithms in upper secondary education: A study of teachers' pedagogical content knowledge. *Computer Science Education*, *33*(1), 61–93. https://doi.org/10.1080/08993408.2021.1935554
Ni, L., Tian, Y., McKlin, T., & Baskin, J. (2023). Who is teaching computer science? Understanding professional identity of American computer science teachers through a national survey. *Computer Science Education*, 1–25. https://doi.org/10.1080/08993408.2023.2195758
Nugent, G., Chen, K., & Soh, L.-K. (2020). The effectiveness of Summer professional development for K-8 computer science teachers. *Society for Information Technology & Teacher Education International Conference*, 86–90. https://www.learntechlib.org/primary/p/215734/
Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9
Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*, *101*(1), 115–135. https://doi.org/10.1037/a0013383
Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002
Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Peteranetz, M. S., Morrow, P. M., & Soh, L.-K. (2020). Development and validation of the computational thinking concepts and skills test. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 926–932. https://doi.org/10.1145/3328778.3366813

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, *2004*(121), 63–73. https://doi.org/10.1002/ir.101

Princeton University. (n.d.). *WordNet*. Retrieved June 2, 2023, from https://wordnet.princeton.edu/

Ravitz, J., Stephenson, C., Parker, K., & Blazevski, J. (2017). Early lessons from evaluation of computer science teacher professional development in Google's CS4HS program. *ACM Transactions on Computing Education*, *17*(4), 16. https://doi.org/10.1145/3077617

Rossi, P. H., Wright, J. D., & Anderson, A. B. (2013). *Handbook of survey research*. Academic Press.

Ruiz, S., Charleer, S., Urretavizcaya, M., Klerkx, J., Fernández-Castro, I., & Duval, E. (2016). Supporting learning by considering emotions: Tracking and visualization a case study. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 254–263. https://doi.org/10.1145/2883851.2883888

Ryoo, J., Flapan, J., Hadad, R., Margolis, J., Amalong, J., Aranguren, L., Campos, E., Knudson, J., Lee, M., & Zuchowicz, M. (2021). Learning with leadership: Perspectives from a statewide research-practice partnership focused on equity-oriented computing professional development for K-12 administrators. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 317–318. https://doi.org/10.1145/3408877.3432569

Sadik, O., & Ottenbreit-Leftwich, A. T. (2023). Understanding U.S. secondary computer science teachers' challenges and needs. *Computer Science Education*, 1–33. https://doi.org/10.1080/08993408.2023.2209474

Sauppé, A., Foley, S. S., Gendreau, T., Hertel, J. T., & Zheng, M. (2019). Building computer science K-12 PLCs in rural communities. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 1275. https://doi.org/10.1145/3287324.3293825

Scott, M. J., & Ghinea, G. (2014). Measuring enrichment: The assembly and validation of an instrument to assess student self-beliefs in CS1. *Proceedings of the Tenth Annual Conference on International Computing Education Research*, 123–130. https://doi.org/10.1145/2632320.2632350

Searle, K. A., Tofel-Grehl, C., Fischback, L., & Hansen, T. (2023). Affordances and limitations of teachers instructional styles when teaching computer science and computational thinking. *Computer Science Education*, *33*(1), 139–161. https://doi.org/10.1080/08993408.2022.2154992

Sentance, S., & Humphreys, S. (2018). Understanding professional learning for computing teachers from the perspective of situated learning. *Computer Science Education*, *28*(4), 345–370. https://doi.org/10.1080/08993408.2018.1525233

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, *2*, 100003. https://doi.org/10.1016/j.nlp.2022.100003

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Shell, D. F., & Soh, L.-K. (2013). Profiles of motivated self-regulation in college computer science courses: Differences in major versus required non-major courses. *Journal of Science Education and Technology*, *22*(6), 899–913. https://doi.org/10.1007/s10956-013-9437-9

Siever, B., & Rogers, M. P. (2019). Micro: Bit magic: Engaging K-12, CS1/2, and Non-majors with IoT & embedded. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 1237–1238. https://doi.org/10.1145/3287324.3287527

Smith, M. (2016, January 30). *Computer science for all*. Whitehouse.Gov. https://obamawhitehouse.archives.gov/blog/2016/01/30/computer-science-all

Stange, M. C. (2020). First bytes: Preparing K-12 educators for CS inclusion. *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 304. https://doi.org/10.1145/3372782.3408110

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson.

Yadav, A., Gretter, S., Hambrusch, S., & Sands, P. (2016). Expanding computer science education in schools: Understanding teacher experiences and challenges. *Computer Science Education*, *26*(4), 235–254. https://doi.org/10.1080/08993408.2016.1257418

Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational thinking in elementary and secondary teacher education. *ACM Transactions on Computing Education*, *14*(1), 1–5. https://doi.org/10.1145/2576872

Zhou, N., Nguyen, H., Fischer, C., Richardson, D., & Warschauer, M. (2020). High school teachers' self-efficacy in teaching computer science. *ACM Transactions on Computing Education*, *20*(3), 1–23. https://doi.org/10.1145/3410631