Affect, Support, and Personal Factors: Multimodal Causal Models of One-on-one Coaching

Lujie Karen Chen
University of Maryland Baltimore County
lujiec@umbc.edu

Joseph Ramsey
Carnegie Mellon University
jdramsey@andrew.cmu.edu

Artur Dubrawski
Carnegie Mellon University
awd@cs.cmu.edu

Human one-on-one coaching involves complex multimodal interactions. Successful coaching requires teachers to closely monitor students' cognitive-affective states and provide support of optimal type, timing, and amount. However, most of the existing human tutoring studies focus primarily on verbal interactions and have yet to incorporate the rich aspects of multimodal cognitive-affective experiences. Meanwhile, the research community lacks principled methods to fully exploit complex multimodal data to uncover the causal relationships between coaching supports, students' cognitive-affective experiences, and their stable individual factors. We explore an analytical framework that is explainable and amenable to incorporating domain knowledge. The proposed framework combines statistical approaches in Sparse Multiple Canonical Correlation, causal discovery, and inference methods for observations. We demonstrate this framework using a multimodal one-on-one math problem-solving coaching dataset collected in naturalistic home environments involving parents and young children. The insights derived from our analyses may inform the design of effective technology-inspired interventions that are personalized and adaptive.

Keywords: multimodal learning analytics, causal discovery, causal inference, parent coaching, affective and cognitive support

1. INTRODUCTION

Studies of human one-on-one tutoring or coaching have a long history (Du Boulay and Luckin, 2016). The insights of effective tutoring strategies and tactics have informed the design the machine-supported tutoring systems such as intelligent tutoring systems (ITS; Graesser et al. 2001). Human tutoring is inherently a rich multimodal interaction process where the students' cognitive-affective experiences and teachers' coaching decisions intertwine. However, most of the existing human tutoring studies focus on the cognitive processes by considering only a single modality of verbal interactions (e.g. Lehman et al. 2012) and rarely incorporate important

signal exchanges via non-verbal channels such as eye gaze or body postures, and the affective experiences.

In recent years, multimodal learning analytics (MMLA; Ochoa 2017) has emerged as a new sub-field of learning analytics that provides methodological frameworks to study a wide range of learning and teaching phenomena. Using MLA, students' certain behaviors can now be tracked and measured using sensors in various contexts, which then may be used to infer students' cognitive and affective states. However, most work is situated in learning contexts when students interact with computer systems rather than with human tutors or coaches. Therefore, there are under-explored opportunities to study human tutoring from the multimodal perspective by leveraging the new sensing technology. On the other hand, we recognize the various challenges in conducting in-depth analyses on a large scale with human tutoring data. The challenges are exacerbated when we move beyond the perceptual level of machine intelligence of recognizing cognitive-affective states and quest for in-depth causal knowledge to be mined from the complex multi-party multimodal datasets. One of this paper's objectives is to fill the gaps in the analytical tools to gain deeper insights into human-human interactions by observing coaching processes. Those tools are building blocks toward the vision of computer-supported tutoring systems that are equipped with higher-level human intelligence in reasoning and decision making.

This paper introduces a novel analytical framework that enables researchers to perform causal inference of multimodal data collected from human one-on-one coaching data. The proposed framework goes beyond traditional human tutoring analysis and attempts to address three additional aspects of human-tutoring data analysis that are under-represented in the current literature:

1. How to leverage rich multimodal data to capture a wide range of behavioral signals observed in the tutoring process;

2. How to discover and characterize the relationship between two intimately related processes, i.e., the coach's decisions and the child's affective-cognitive experiences;

3. How to uncover causal relationships among the descriptors from multimodal dyadic interaction data.

Achieving these goals requires us to resolve the tension between the complexity of the multimodal data streams that are often high-dimensional as well as observed at high frequencies and the desire for explainable and transparent modeling to support interpretable causal discovery and inference. The proposed pipeline is a two-step procedure that combines the sparsity-induced search for systems of multiple composite variables with high correlation, followed by causal discovery and inference.

We demonstrate the framework using a dataset collected from 15 parent-child dyads where parents coach their children on challenging math problem-solving at home. We frame the coaching process as an implicit optimization process by a parent to resolve the *assistance dilemma* (Koedinger and Aleven, 2007). Specifically, on a moment-by-moment basis, parents need to render support to induce the right amount of productive struggle (Kapur, 2014). In essence, parent coaches need to strike a delicate balance between coaching support that is too limited, leading to an overly frustrating experience, and support that is too intense, which may degrade the perseverance opportunity and lead to an undesirable child's state of "learned helplessness" (Miller and Norman, 1979). Concretely, parents need to make decisions on the *amount*, *type*

(cognitive, meta-cognitive, or social/emotional), and *timing* of their support. In this paper, we are interested in understanding parents' coaching decisions regarding the amount and type of supports that we explore in the context of children's cognitive-affective experiences, which can be observed and estimated from multimodal behavioral signals. Additionally, we study how the interactions are influenced by a child's stable traits such as personality or belief. We are interested in recognizing **what** kind of intervention has happened. More importantly, we seek to understand **why** certain types of intervention take place and what are the consequences of those interventions. We hope to recover those intricate pathways of influence from the observational data we gathered in the multi-subject study and, where appropriate, from domain knowledge. The methodology we employ is a combination of structural model discovery and statistical hypothesis testing.

The analyses described in this paper are based primarily on ground truth obtained through manual annotations of the multimodal dataset. To explore the potential of the proposed approach to scale up with automatic annotation, we investigated machine learning models (Goswami et al., 2020) to derive automatic recognizers for those annotations in an independent line of work, laying the foundation for future efforts towards large-scale multimodal human tutoring studies. It is, however, not in the scope of this paper to discuss details of those models.

In the remainder of the paper, we will first overview related work in math education, multimodal learning analytics, and causal inference in educational applications. We will then introduce the dataset in Section 3 and highlight a few key measures to be considered in the downstream analysis, followed by an overview of the analysis pipeline in Section 4. Section 5 will then introduce the application of multiple canonical correlation analysis (multiple CCA, or mCCA) and report results obtained with it. In Section 6, we will detail the procedure for causal discovery and inference using the output from mCCA. To scaffold the complex analysis, we will first introduce a simplified model with only four factors. We will then present results from an enriched model with seven factors. We will conclude the paper by summarizing the findings, discussing their implication and future directions of research.

2. Related Work

2.1. Cognitive-affective states during math problem solving among young children

Different from math practice, which consists of routine exercises (i.e., problem-as-exercise), math problems come without immediate solutions (i.e., problem-as-problematic; Schoenfeld 2016). This distinction leads to potentially very different cognitive-affective experiences for students. Due to the inherent uncertainty in reaching a solution, authentic math problem solving often sends students onto a ride of *emotional roller-coaster* (Chen et al., 2016), wavering between positive affect when progressing smoothly (i.e., Engagement Concentration or EC; Baker et al. 2010), and negative confusion or frustration states when the progress is obstructed (i.e., Cognitive Disequilibrium or CD; D'Mello et al. 2012).

Studies on affect dynamics show that while productive struggles can induce deeper learning (Kapur, 2014) , unresolved issues may lead to disengagement or boredom and eventually diminish learning outcomes (D'Mello and Graesser, 2012). It is then crucial to engineer the learning environment to maximize the productive struggle while minimizing the unproductive effects. This can be especially challenging with young children whose self-regulation skills

(SRL; Zimmerman 2000) are still being developed (Di Leo et al., 2019).

In recent years, a thread of work has emerged with a specific focus on studying math problem-solving induced cognitive-affective experience with young age groups, beginning to shed light on that challenging issue. For example, Di Leo et al. (2019) studied young students in 5th and 6th grades engaged in independent math problem-solving. They illustrate the pathway from students' belief to affective experience, problem-solving strategy, and problem-solving experience. These findings have motivated a follow-up intervention study by the same authors (Di Leo and Muis, 2020) who leveraged explicit instructions in problem-solving strategies. We study math problem solving with similarly aged children; however, we take an alternative approach. Instead of focusing on a child's self-regulation in the face of obstacles, we study how assisted regulation provided by parent coaches may play out amid children's cognitive-affective states in the process. Similar to Di Leo et al. (2019), we also measure the aspects of personal traits that may serve as the antecedent of academic emotion as predicted by the control-value theory (Pekrun and Stephens, 2010). In our study, control is approximated by children's belief in their math ability, while children's self-reported math interest indicates value.

From a methodological point of view, our work differs from Di Leo et al. (2019) in two critical dimensions. Firstly, in our study, the cognitive-affective states are objectively observed. This practice is different from Di Leo et al., which comprises retrospective self-reporting or verbal coding of students' think-aloud transcripts. Secondly, we adopt a data-driven causal discovery framework while incorporating domain knowledge instead of solely relying on theory in determining the causal structures, as implied in their path analysis method.

2.2. Multimodal Learning Analytics

Multimodal learning analytics (MMLA) is an emerging multidisciplinary research area that integrates learning science, affective computing, and human-computer interaction (Cukurova et al., 2020). It leverages modern sensing technology and computational advances to analyze complex human behavior and holds the potential to render a detailed and holistic picture of the learning processes (Drachsler and Schneider, 2018).

As noted from recent reviews by Sharma and Giannakos (2020) and Mu et al. (2020), most of the existing research pertains to the online learning context where students interact with computer systems such as intelligent tutoring systems, e.g., Hutt et al. (2019), or educational games, e.g., Giannakos et al. (2019) and Emerson et al. (2020). There is a relatively small amount of work situated in offline or physical spaces. For example, Worsley and Blikstein (2018) used MMLA to explore learning in maker space when students work with physical objects. Zhu et al. (2019) studied the cognitive and emotional dynamics of elementary school students in physical classrooms. Additionally, we note that most of the work models students (either as an individual or as part of a group, as in Martinez-Maldonado et al. 2019); however, teachers or coaches are rarely the modeling targets. Exceptions include work from Prieto et al. (2016) or Prieto et al. (2018), where the objective is to model how a teacher manages a classroom. Besides, as noted in Sharma and Giannakos (2020), most of the research has taken place in a lab environment with college students predominately from the subject pool. It might be because data collection from ecologically valid environments such as home or school could be challenging and complicated. Our work leverages a unique multimodal one-on-one coaching dataset collected from the naturalistic home environments involving parent coaches and young children. This dataset renders a rare opportunity to explore the multimodal interactions between children and their parent

coaches, which is largely missing from the current MMLA literature.

This paper's second unique feature is the exploration of methods for the downstream analysis of teachers' decisions given students' cognitive-affected states, recognized from the *detection layers*. The goal of this analysis is different from the majority of educational affect-related multimodal work reviewed in D'Mello and Kory (2015) and Yadegaridehkordi et al. (2019), in which detection is the main objective. There are only a few lines of work that aim to achieve similar goals as we do; however, those studies involve students in online learning environments. For example, as described in Grawemeyer et al. (2017), the authors explicitly modeled a *reasoning layer* to infer the type of feedback that teachers tend to give based on the student affect estimation. The inference model uses data from Wizard-of-Oz experiments[1] when students interact with a fraction tutoring system. Another example is from Santos et al. (2014), where a model is learned from educators who responded to students' affect dynamics while a student is working with an online learning tool. In this observational study (Porayska-Pomsta et al., 2008), the tutors worked with students in an online learning environment via a text chat interface. The tutors were asked to reflect on how students' contextual factors (e.g., confidence, interest, and effort) may have contributed to their tutoring actions. That information was gathered through tutors' think-alouds and post-hoc walkthroughs. In our study, we don't have access to tutors' cognitive process in decision making; instead, we rely on causal inference from empirical observations data to illuminate the plausible relationship between students' affect and tutors' coaching decisions.

Overall, our work can be viewed as a particular case of model-based discovery (Baker and Yacef, 2009); specifically, model-based *causal* discovery as pointed out in Fancsali (2015). Our main objective is to discover the causal relationship between parents' coaching decisions and their children's cognitive-affective states.

2.3. Causal Inference from Observational Data in Education Applications

The ultimate goal of educational research is to identify potentially useful interventions to improve educational outcomes. While the randomized control trial (RCT) is the gold standard methodology to uncover causal knowledge, it is often expensive, time-consuming, and sometimes impractical or unethical to implement with students in the real world. Causal inference from observational data provides an alternative that allows us to exploit even large amounts of empirical data in the search for answers beyond apparent correlations. However, its adoption in the educational data mining or learning analytics community is still rather limited. For example, using a discovery and inference framework, Fancsali (2015) explored the confounding role of carelessness in explaining the counter-intuitive relationship between affective states of confusion/boredom and learning outcomes by mining a dataset of student interactions with the Algebra cognitive tutor. In a different paper with a similar dataset, the authors elucidate the causal relationship between prior knowledge, affective experiences, gaming behaviors, and learning outcome via the framework of *causal discovery with a model* (Fancsali, 2014). With data gathered from an online learning environment, Koedinger et al. (2016) analyzed the student interaction log data. They demonstrated the causal effect of active engagement on learning outcomes, using a causal discovery and inference toolkit TETRAD[2], which we also adopt as part of

[1] In those experiments, students interact with a system that is controlled by teachers. As such, what is effectively being modeled is the human teachers behind the scenes.

[2] https://github.com/cmu-phil/tetrad

our pipeline. de Carvalho et al. (2018) conducted causal inference on students' online behavior patterns using a different toolkit, GeNIe[3], using data from a learning management system. Besides, we see the applications of causal inference to discover knowledge dependencies (Scheines et al., 2014).

3. Dataset

We demonstrate the proposed analytical framework using a multimodal one-on-one coaching dataset collected in naturalistic home environments. In this section, we first give an overview of the study protocol and data collection procedures. We then introduce a few critical preprocessing steps, including the annotations of child participants' cognitive-affective states and parent coaches' support types. We end the section with a description of the session-level dataset compiled for the subsequent analyses.

3.1. Data Collection

With the Institute Research Board's approval at Carnegie Mellon University, we recruited 15 parent-child dyads from the local community. Child participants were eight to twelve years old (equivalently: third to sixth grades) whose parents were interested in math problem-solving coaching at home. Two-thirds of the children were in 4th or 5th grades at the time of data collection, with an almost even gender distribution. Regarding ethnic background, only 1 out of 15 child-parent dyads was from African American family with the rest coming from Caucasian (n=9) or Asian (n=5) backgrounds. All parents held at least bachelor's degrees from diverse disciplinary backgrounds, including professional writing, biology, humanity, business, management, computer science, and engineering. Three of them hold Ph.D. degrees, with one parent having a doctoral degree in mathematics.

In each session, the child worked through one problem while thinking aloud [4]as much as feasible, and his or her parent provided support as needed. We supplied parents with a pool of math problem-solving resources[5] from which they chose what based on their subjective assessment might pose a sufficient challenge for their children. We made this decision as we were concerned that individual children may have varied problem-solving experiences and levels of frustration tolerance. As such, parents may be better informed than researchers, given the amount of their implicit knowledge of their child's capabilities, including cognitive skills and affective responses that are often not easily accessible to outsiders.

We collected data from 76 sessions with a cumulative duration of about 624 minutes, or 10.4 hours, and a mean duration of 8.2 minutes per session. The shortest session was less than 1 minute, while the longest one was about 22 minutes long. We collected audio/video recordings of these sessions taken at home by parents using consumer-grade recording devices such as webcams or smartphones. Specifically, we collected audio streams of both parent and child participants and frontal view video streams only of child participants. In addition to the audio/video

[3]https://www.bayesfusion.com/

[4]Before the recording started, we provided training to parent and child on thinking aloud. During the session, parents were asked to remind, but not force, their child to think aloud. Often, we observed that a child would stop thinking aloud if he or she was engaged in deep thinking.

[5]The resources include problem sets from math competitions for elementary and middle students, such as MOEMS (https://www.moems.org/), Math Kangaroo (http://www.mathkangaroo.us), MATHCOUNTS (www.mathcounts.org), or AMC series (https://www.maa.org/math-competitions)

data, we also asked parents to complete questionnaires describing their subjective assessment of the child's experience during and after each session (Appendix A). Those assessments mainly reflect the academic affect commonly mentioned in the literature, including frustration, confusion, joy, surprise, etc. In addition, we also collected child participants' responses to survey instruments on achievement/goal (Elliot and Murayama, 2008), math interest (Linnenbrink-Garcia et al., 2010), self control (Tsukayama et al., 2013), self-efficacy (Bandura, 2006), grit (Duckworth and Quinn, 2009), effort regulation (Pintrich et al., 1991), help-seeking (Pintrich et al., 1991) and personality (John and Srivastava, 1999).

3.2. Data Preprocessing and Annotation

Multimodal interaction data is inherently rich and notoriously challenging to analyze. Its practical use requires preprocessing and annotation steps to support the goals of analysis. In our case, we aim to elucidate the relationship between a child's cognitive-affective experience and parents' support decisions. The annotation work was carried out by the first author and her research assistants.

We annotated the voice activity from video recordings of parent-child interactions at the utterance level. Those annotations explicate "who talks when." All the utterances were further transcribed. In addition, we annotated each child's eye gaze toward their parent. Besides those, we have also implemented additional types of annotations and featurization unique to our study, as described below.

Annotation of Cognitive-Affective States. We annotated the apparent child participants' cognitive-affective states for each 10s video segment[6] throughout each session, using multimodal behavioral cues from visual channels such as gross body movements and facial expressions, and from verbal channels such as talking speed, dis-fluency (e.g., "uhm" utterances), as well as energy and loudness. We specifically annotate data for three different types of cognitive-affective states (see Appendix C.1 for annotation details):

- Cognitive Disequilibrium (CD; D'Mello et al. 2012): A state of confusion, frustration, indecisiveness, or struggle in the face of an impasse. It may occur during any problem-solving stage, including problem understanding, planning, or implementation;
- Engagement Concentration (EC; Baker et al. 2010): A state suggesting a smooth progression in problem solving. Compared with the flow experience described in Csikszentmihalyi (2013), those states observed with young children in our study are more likely characterized by lower intensity or shorter duration;
- Neutral: A state with no strong indications of either CD or EC.

To evaluate inter-rater reliability, we randomly selected one session from each subject for which annotations from two independent raters were collected. We then compared the codes and resolved any discrepancies after discussion. For videos with large discrepancies, raters

[6]The decision of using 10s video segment as the unit of analysis was largely informed by the "thin-slice" (Ambady and Rosenthal, 1992) approach. From our empirical observation, video segments of this duration tend to give us sufficient information to understand the context and evaluate the affective state of the child but are not so long that it becomes difficult to capture fine-grained affect dynamics.

recoded the videos from scratch until a certain level of consensus was reached. Then the same two raters coded a second video from the same subject until two raters converged. As a result, around 20% of the total number of segments in the complete dataset were coded by two raters. The overall final inter-rater reliability (measured with Cohen's kappa) is 0.51 [95% CI: 0.46, 0.57] for three-class-category (CD, Neutral and EC); two-class kappa (CD vs. Non-CD) is 0.58 [0.51, 0.66]; and two-class kappa (EC vs. Non-EC) is 0.61 [0.54, 0.58]. The observed moderate inter-rater consistency of collected annotations suggests that it is not unreasonable to use these labels in the subsequent analyses as a somewhat noisy proxy for ground truth. The coding task for the rest of the videos was then split between the same two raters.

In a separate study (Goswami et al., 2020), we explored machine learning models to automatically discriminate between CD and EC segments using a semi-supervised framework based on the low-level signals extracted from audio/video streams. Those models achieved reasonable predictive accuracy measured via cross-validation on the manually labeled subsets of data. The model achieved an area under curve (AUC) score of about 0.80. When further validated, those automatic recognition methods have the promise to scale up our methodology to future large datasets.

RESERVE CHART FEATURES. Reserve charts summarize the moment-by-moment temporal evolution of a child's cognitive-affective experience during a given problem-solving session. These charts are derived from the cognitive-affective state annotations described above. Statistics such as current values, trends, or cumulative level of the states estimated at any point in the session can be used as features in downstream analysis to characterize children's cognitive-affective experiences.

Figure 1 shows an example reserve chart. The top panel tracks the state ($EC = 1, CD = -1, Neutral = 0$) for each 10s segment of a session. The bottom panel depicts the cumulative sum of these state values. We view it as a proxy for the temporal evolution of the "reserve level" of a child's psychological resources while they are progressing through problem-solving exercises. The reserve level initializes at zero and will trend up with continuous exposure to positive experience from smooth progression (EC) while trending down with sustained negative experience of struggle (CD). Neutral states will not alter the reserve levels. Visually, a straight downward path, reflective of a prolonged stretch of struggles, would warrant close monitoring as the child is likely to run the risk of depleting their reserve soon, thus justifying an immediate intervention. On the other hand, if a child is progressing smoothly without much trouble (as manifested by an upward trend), fewer monitoring resources may be necessary.

ANNOTATIONS OF THE TYPE OF SUPPORT. We annotated parents' support types based on audio streams of parents' utterances. In annotation, we refer to the audio/video segments around the target utterance. In doing so, we implicitly consider multimodal signals including parents' para-linguistic features while talking such as loudness, tone, and voice pitch. Contextual information of child participants' verbal or non-verbal responses may also play a role in the annotation process.

A high-level taxonomy of parents' support types is (see Appendix C.2 for annotation details):

- *Cognitive Support*: Direct and targeted support in the form of explicit coaching statements, including asking questions or providing explanations that are specific to a given problem;

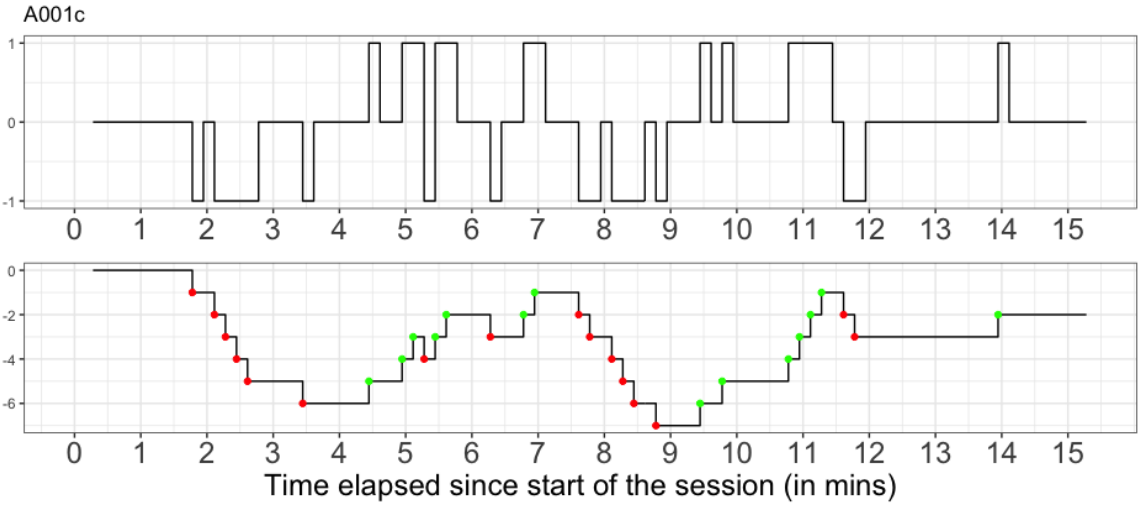

Figure 1: A reserve chart for an example session A001c. The top panel displays the time series of annotated cognitive-affective states to which we assign numeric values: Cognitive Disequilibrium (CD, value=-1), Neutral (value=0), and Engaged Concentration (EC, value=1). The bottom panel shows the cumulative sum of the numeric values of the states from the top panel, assuming initialization at 0 at the start of the session.

- *Meta-cognitive Support*: Hints on general problem-solving strategies such as drawing a picture or making a list, often without detailed instruction or guidance as in case of *Cognitive Support*.

- *Social/Emotional Support*: Support for boosting a child's self-efficacy. Parents may use language to provide assurance (e.g., by saying "it's okay to struggle"), communicate growth mindset messages (Dweck, 2008), praise effort (e.g., "you've been working so hard"), or encourage (e.g., "you can do it!" or "keep trying"). Where applicable, we attached a modifier of "+" or "-" to the labels. For example, S+ indicates positive social/emotional support, while S- denotes a negative alternative. While positive social/emotional supports use encouraging phrases such as "you are doing great" or "why not try a little bit", negative supports may involve discouraging messages that may threaten self-efficacy, examples including "this should not be a very hard problem" or "now does it make any sense?"

About 20% of parents' utterances are randomly selected to be annotated by two independent raters. The overall inter-rater reliability (Cohen's kappa) for multivariate labels is 0.74 [95%CI 0.68, 0.78] (with modifiers) and 0.86 [0.81, 0.90] (without modifiers). Kappa for binary labels are 0.91 [0.88, 0.95] for *Cognitive Supports* (66% of annotations), 0.82 [0.68, 0.94] for *Meta-Cognitive Supports* (6% of annotations), and 0.65 [0.47, 0.82] for *Social/Emotional Supports* (4% of annotations). The observed relatively high inter-rater reliability scores suggest consistency of the resulting labels appropriate for downstream analyses that depend on those annotations.

3.3. Session Level Dataset

The analysis is based on the session-level summary statistics from 53 sessions.[7] In this dataset, each instance represents one session. We compiled a list of 59 features characterizing these sessions, organized into seven groups summarized below. We present a detailed description of all the 59 variables in Appendix C.

1. Child's affective-cognitive experience

    - *Affect*: Metrics describing the aggregated amount and mixture of CD/EC/Neutral of each session;
    - *Stress*: Metrics describing the dynamics of reserve chart (Figure 1) highlighting extreme points of the chart (peaks and valleys) and streaks of continuous "up" (EC) or "down" (CD) episodes.

    While *Affect* describes the overall mixture of affective experience, *Stress* summarizes the details of the fine-grained process of a child's affective-cognitive experience.

2. Parent's support

    - *Support*: Metrics describing the types of support with regard to cognitive, metacognitive, and social/emotional types;
    - *Interact*: Metrics describing the interaction patterns between parent and child, including voice activities ("who talks when") and eye gaze patterns. We may view those metrics as proxies of the amount of support.

    We may interpret *Support* and *Interact* along two different dimensions of parents' support in terms of the *type* and *amount* respectively.

3. *Assess*: Per-session assessment of child's experience during and at the end of the session, completed by the parent at the end of each session (see Appendix A).

4. *Profile*: Per-child measurements from survey instruments and questionnaires completed by the child. (see Appendix B).

5. *Tot*: Time-on-task, or duration of the session.

4. Analysis Pipeline

Figure 2 depicts the analysis pipeline designed to automatically or semi-automatically extract and identify structural and causal relationships from high-dimensional human-to-human interaction data by combining data streams with varying temporal resolutions from multiple modalities.

The pipeline inputs include data streams with high temporal resolution, such as those describing child participants' affective-cognitive processes, parents' moment-by-moment supports,

[7] From the total of 77 sessions observed, we leave out sessions without complete information. Common instances of incomplete information arose when the parents forgot to answer the post-session questionnaire on the child's affect evaluation. Because the data was collected at home, it is hard to recover this information after the data was collected.

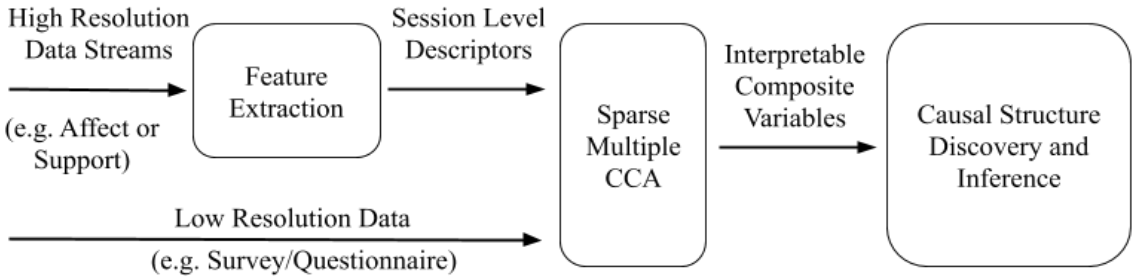

Figure 2: Analysis pipeline designed to support discovery of causal insights from multimodal interaction data.

and the interaction dynamics between parent and child. We extracted session-level summary statistics or descriptors from those data streams, combined with data elements with low temporal resolution data such as the elements collected from per-session logs or per-subject survey responses. We organized the resulting high-dimensional feature set into multiple groups of features according to their roles in the coaching processes, for example, with respect to child's affect, parents' supports, or child's profile or stable traits. These groups of variables are used as inputs to the Sparse Multiple CCA procedure (Witten and Tibshirani, 2009) to learn a sparse representation for each group by simultaneously maximizing the correlations among the groups of multiple features. The resulting composite variables, represented by linear combinations of small sets of features, are then used as inputs for causal structural discovery and inference. Section 5 briefly explains the sparse multiple CCA method and presents relevant results. Section 6 details the procedures and results from causal structure discovery and inference.

5. Sparse Multiple CCA

The standard canonical correlation analysis (CCA) takes as inputs two matrices of numbers, $X_1 \in R^{n \times p_1}$ and $X_2 \in R^{n \times p_2}$, each comprised of a set of features in dimensions p_1 and p_2 respectively, represented on the same set of n data samples. The goal of CCA is to find linear combinations of variables in each dataset that yield high linear correlations between features in X_1 and X_2. The core idea is similar to that of principal component analysis (PCA) in that both methods produce linear combinations of the original features of data, with the exception that CCA's "principal components" consist of pairs of such linear combinations – one for each of the data sets – while PCA only yields single combination per component, since it only works with one data set. When we project the original data on the resulting composite dimensions defined by these linear combinations, we will notice that while the principal components are selected by PCA to maximize variance of data after such projections, CCA selects projections that maximize correlations between data sets.

Sparse multiple CCA or sparse mCCA (mCCA; Witten and Tibshirani 2009) extends the standard CCA by imposing sparsity constraints. This yields CCA projections that rely on subsets of all features in X_1 and X_2, and therefore to compact, more interpretable representations of data. This capability makes mCCA a good candidate for discovering meaningful multiple-to-multiple correlations between datasets of high dimensionalities compared with the number of samples. Another benefit of sparsity is the improved transparency and interpretability of the

resulting composite variables. This benefit results from the L1-norm regularization procedure that effectively shrinks many of the feature weights to zero. An additional useful extension of mCCA allows the optimization algorithm to operate on multiple datasets X_i ($i = 1 \ldots K$), instead of just two at a time, as in the standard CCA setting. As a result, for each set of features, we can devise a compact composite variable to represent it. The composition of this variable will be optimized to maximize its correlation with all other such composite variables devised simultaneously for all other feature sets. Formally, the goal of mCCA is to find vectors of feature weights $W_i, i = 1 \ldots K$ that optimize:

$$max_{W_1 \ldots W_K} \sum_{i<j} W_i^T X_i^T X_j W_j \text{ subject to} \|W_i\| \leq 1, L_1(W_i) \leq c_i \forall i \quad (1)$$

where L_1 is the regularization penalty function in the L1 norm form, and c_i are hyperparameters that upper-bound W_is. The higher the upper bounds, the more complex models the algorithm will be willing to explore. It is assumed that $1 \leq c_i \leq \sqrt{p_i}$, where p_i is the dimensionality of the *ith* dataset.

Both the ability to jointly handle multiple datasets and to induce sparse representations are relevant properties of the mCCA approach in support of our analysis goals. Our dataset comprises descriptors grouped into several distinct categories. Each category describes a different aspect of the coaching process or specific factors that may influence the processes, such as child participants' affective-cognitive experience and stable traits as well as parents' support characterization. To efficiently discover causal relationships among those factors, we hope to find a representation for each group of features that would maximize the overall correlations, given that correlation is often a prerequisite for causation. mCCA helps to achieve and maintain such focus. Moreover, with 59 features and 53 instances of data, we simply need to encourage sparse models to mitigate the risks of over-fitting. Further, the learned sparse weights of the original features of the data allow us to assign meaning to the composite variables, which enhances the interpretability of downstream analysis for causal structural discovery and inference. This improved transparency not only makes it easier to incorporate domain knowledge, but it may also further facilitate the model understanding and critique, which are essential steps toward human-centered analytics, very often beneficial in various contexts of practical application of artificial intelligence.

We fit mCCA using R package PMA[8]. The data was first standardized per feature to zero mean and unit standard deviation before applying mCCA. We then performed grid search for optimal values of c_i to find a model with a reasonable total correlation and compactness. Figure 3 plots the average correlation (total correlation (see Equation 1) normalized by the number of pairs of correlating CCA components) as a function of c_i. The overall model complexity (approximated by the average number of non-zero weights) is printed as a digit on top of the curve. This plot shows an overall trend of increasing correlation with the increased allowed model complexity. For the sake of interpretability, however, we prefer simple models capturing a reasonable amount of correlation in data, even if the captured correlation is slightly lower than the attainable maximum. The dotted line corresponds to a penalty value of 1.4 (i.e., $c_i = 1.4$, $\forall i$), which leads to a model with an average of 3 variables per group. In the following analysis, we will use composite variables derived from this specific model.

Table 1 lists the composite variables and the underlying features and weights derived using

[8]https://cran.r-project.org/web/packages/PMA/index.html

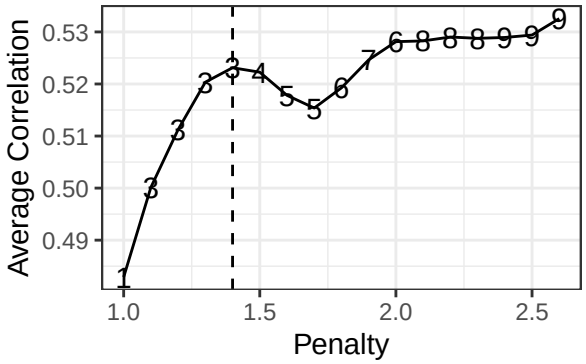

Figure 3: Grid search for the optimal setting of the mCCA regularization parameter. The plot shows model performance, measured as the average correlation of returned CCA projections, as a function of penalty upper bounds c_i. Labels on the plotted line reflect the average number of non-zero weights across all involved feature groups. The dotted line reflects the chosen compact model with sparse weights and reasonable overall correlation.

the mCCA approach. Each of the composite variables is a linear combination of features selected from a specific feature set. By enforcing sparsity of the resulting mCCA models, each composite variable only uses a few original data features which had been assigned non-zero weights. An example composite variable is $\textit{Support} = 0.49 \cdot \textit{C\_count} + 0.87 \cdot \textit{C\_length} + 0.04 \cdot \textit{CM\_count}$. In this equation, *C_count* is total counts of cognitive supports observed during a given session, *C_length* is the cumulative duration of cognitive supports, and *CM_count* is the total counts of cognitive and meta-cognitive supports combined.

In Table 1, we also provide plausible interpretations for the high value of a composite variable by accounting for the weights and their signs in the context of underlying original features. For example, when interpreting the composite variable *Support*, we note that most of the underlying variables are related to cognitive support. Since the signs of the weights are all positive, we pose that a high value of *Support* should be associated with a high level of cognitive support, and vice versa. As another example, the value of the composite variable *Stress* may be reduced by increasing the value of e.g. *max_EC_length* (i.e., the longest EC streaks) as well as other original features selected to form this composite variable, which collectively signal struggle experienced by the child during one of their problem-solving session.

In some cases, however, the assignment of meaning is not obvious when the signs of weights fail to render a coherent interpretation of the composite variables in question. For example, composite variable *Profile* is defined as $\textit{Profile} = 0.20 \cdot \textit{Grit} - 0.95 \cdot \textit{Extrovert} - 0.26 \cdot \textit{Self-efficacy}$. Those weights suggest that a child with a high *Profile* score is likely to have a high grit score, with an introvert personality type and low self-efficacy. It is unclear how we could characterize a child with this specific profile. Therefore, we leave the interpretation of this composite variable as undefined and will consult the underlying variables in downstream analysis as necessary.

Figure 4 shows the scatter-plot matrix of composite variables after applying sparse weights to the original feature sets as shown in Table 1 to each sample in the dataset. The diagonal plots depict density distributions of each feature set data projected onto their composite variable. The below-diagonal plots show scatters of data projected on pairs of composite variables corresponding to the feature sets in the respective rows and columns of the matrix plot. The

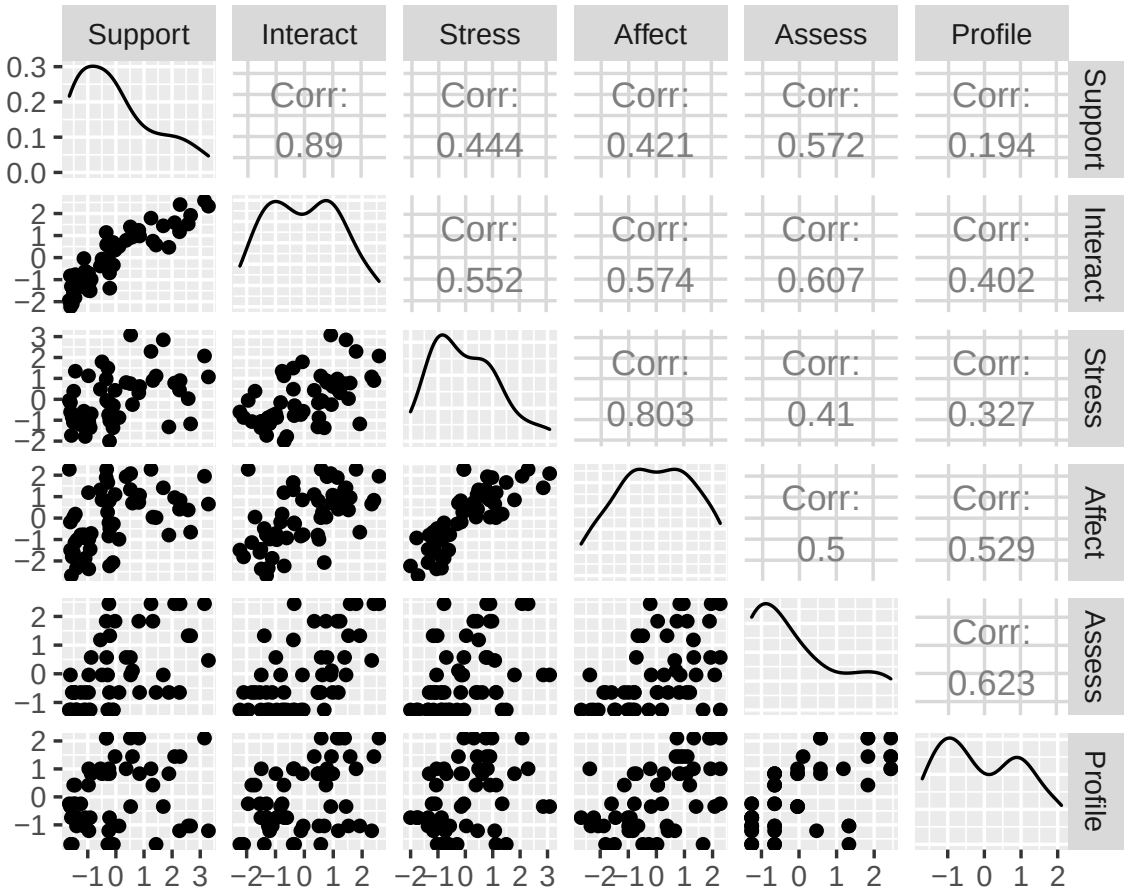

Figure 4: Scatter-plot matrix of the composite variables yielded by the Sparse Multiple CCA model.

Table 1: Composite variables (first column) with associated underlying original features (second and third column) and weights (fourth column), estimated using the Sparse Multiple CCA model. Each composite variable is a linear combination of the underlying features. Please refer to Appendix B and C for feature name description.

Composite Variable	**High Value Means**	**Feature Name**	**Weights**
Support	More cognitive support	C_count	0.49
		C_length	0.87
		CM_count	0.04
Interact	More parent talk	Parent_talk_count	0.17
		Parent_talk_duration	0.29
		Parent_talk_pctg	0.94
Stress	More struggle	Global_low	-0.93
		Global_high	-0.23
		Max_EC_length	-0.28
		Num_EC	-0.06
Affect	More struggle	Pctg_EC	-0.95
		Pctg_CD2	0.23
		Pctg_EC2	-0.23
Assess	High frustration/ confusion	During_frustrated	0.90
		End_accomplished	-0.06
		End_confused	0.44
Profile	Undefined	G(Grit)	0.20
		PE(Extrovert)	-0.95
		SE(Self-efficacy)	-0.26

above-diagonal cells show the values of linear correlation coefficients computed from the scatter plots in the corresponding cells located symmetrically across the diagonal of the matrix plot. As can be seen, the pair of composite variables *Support* and *Interact* exhibits a relatively high correlation, which does not come as a surprise since they describe two related but complementary aspects of parents' support decisions. While *Support* reflects the mix of the types of support given (e.g., the proportion of cognitive support), *Interact* is mainly concerned with the amount of support as approximated by the number of parent talk episodes. We also observe a high correlation between *Stress* and *Affect*, presumably because they originate from the same underlying affective-cognitive processes of child participants. Permutation test of significance (n=1000) shows that for all but one pair (*Support* and *Profile*), the observed correlations are statistically significant with p-values < 0.01.

Figure 5 shows the 2-dimensional multi-dimensional scaling embedding of the considered composite variables using the complements of pairwise correlations as the distance metric ($1 - r_{ij}$, where r_{ij} is the observed linear correlation coefficient between composite variables i and j). In this plot, composite variables that correlate well are projected in each other's vicinity and vice versa. It is not surprising to note that the affect-related variables (i.e., *Affect* and *Stress*) are grouped together, as are the support-related variables (i.e., *Interact* and *Support*).

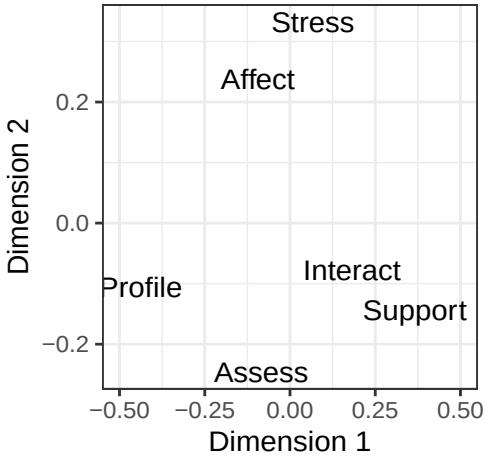

Figure 5: Multi-dimensional scaling plot using the complement of pairwise correlation as a distance metric.

6. Causal Structural Discovery and Inference

We will now discuss methods to explore the casual relationships among multiple composite variables constructed using the mCCA approach described above. We compiled a dataset of 53 instances (as before, one per session), each represented by a vector of the six composite variables, with an additional scalar variable *Tot* ("Time-on-task"). This variable is believed to be highly correlated with the individualized calibration of the level of difficulty of the problem. Presumably, the more difficult a problem is for a given child, the longer they may need to work on it, and the more likely the parent will need to intervene, which would then stretch the session longer.

At a high level, we are interested in understanding the multivariate causal relationships among children's affective-cognitive experience, parents' support, and other factors such as children's profile variables. To accomplish that, we present two sets of analyses. The first relies on a model with only four factors, while the other uses all seven factors.

Analysis of the 4-Factor Model The input to the 4-factor model includes *Affect*, *Support*, *Tot*, and *Profile* which are representative of four main types of session-level measures captured in this study. Except for *Tot*, other variables are composite variables derived from data using the mCCA approach as shown in Table 1.

The analysis starts with a search for the most probable causal graph pattern (i.e., equivalence class of graphs or ECG). We then derive a few variations of directed acyclic graphs (DAGs) from the found causal graph patterns by incorporating domain knowledge. Each DAG represents a competing hypothesis about true causal relationships among a set of variables. Those DAGs are then used as inputs to estimate linear Gaussian Structural Equation Models (SEMs). We evaluate the goodness-of-fit of those alternative models to underlying data to assess their plausibility. All procedures, including causal model discovery, model specification, and estimation, as well as statistical testing of goodness-of-fit, were carried out using TETRAD toolkit version 6.8.0-0.[9]

[9]http://www.phil.cmu.edu/tetrad/

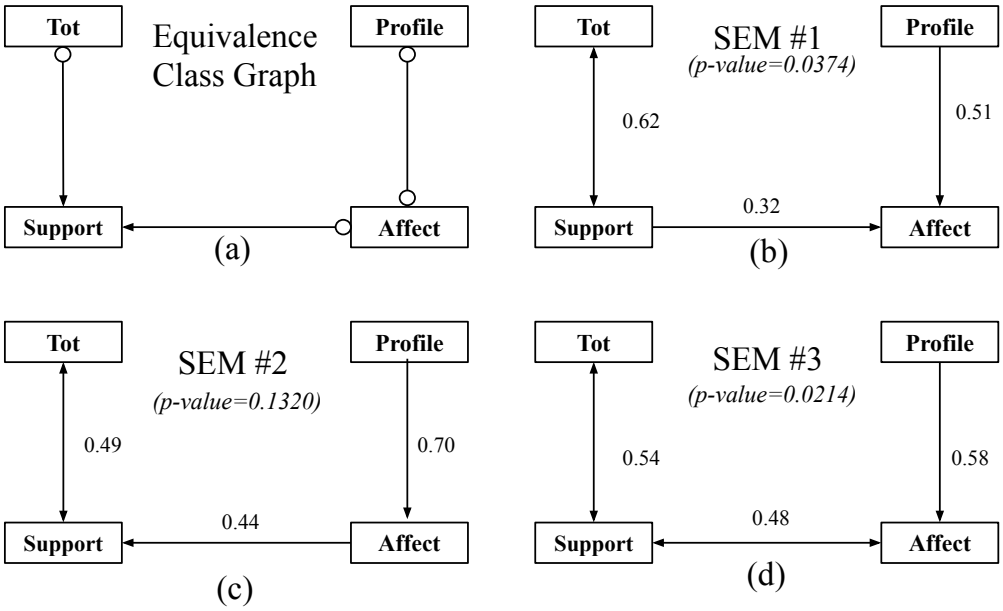

Figure 6: Causal models for the 4-factor model. Subplot (a) represents the equivalence class graph discovered by the GFCI search algorithm implemented in TETRAD; Subplots (b), (c), and (d) are three alternative DAGs derived from (a). Numeric values on the directed edges denote coefficient estimates, and those on double-headed arrows reflect correlations of error terms. Model p-value is from the goodness-of-fit χ^2 test.

Figure 6 presents the results of causal graph search and model estimations. Subplot (a) is the equivalence class graph discovered by TETRAD using the GFCI algorithm with Fisher Z-test score and SEM BIC score as objectives. The interpretation of the graph structure shown in Figure 6(a) is as follows:

- *Tot* and *Support*: either *Tot* is a direct cause of *Support* or there is an unmeasured confounder between those two variables, or both;
- *Affect* and *Support*: either *Affect* is a direct cause of *Support* or there is an unmeasured confounder between those two variables, or both;
- *Profile* and *Affect*: either (i) *Profile* is a cause of *Affect* or (ii) *Affect* is a cause of *Profile* or (ii) there is unmeasured confounder between the two variables or (i) and (iii) or (ii) and (iii);
- Missing links: there is no direct causal relationship between the disconnected variables. For example, there is no edge between *Profile* and *Support*, which suggests that *Profile* has no direct influence on parent's decisions regarding rendering support.

The graph in Figure 6(a) does not fully specify causal relationships. The other three graphs in Figure 6 use it as a starting point and make specific hypothetical assignments based on the application of domain-specific intuition. For example, for the pair of variables *Tot* and *Support*, we believe that there is a possible unmeasured confounding variable related to the problem

Table 2: Goodness-of-fit testing statistics for three alternative structural equation models corresponding to graph structures in Figure 6(b), (c), and (d).

Model	SEM #1	SEM #2	SEM #3
Degrees of Freedom	3	3	3
χ^2	8.4580	5.6140	9.6886
P-Value	0.0374	0.1320	0.0214
BIC Score	-3.4530	-6.2970	-2.2230

difficulty level as calibrated to individual child's ability. As such, we assign a double-headed arrow link between those two variables to encode this belief. For the pair of *Profile* and *Affect*, we believe *Profile* is an exogenous variable that can only be the cause but not the effect. As such, we designated a directional arrow edge leading from *Profile* to *Affect*. Regarding the relationship between *Support* and *Affect*, we are interested in testing the following three alternative hypotheses:

- H1 as in subplot 6(b): *Support* influences *Affect*, yielding structural equivalence model SEM #1;
- H2 as in subplot 6(c): *Affect* influences *Support*, SEM #2;
- H3 as in subplot 6(d): *Affect* and *Support* are influenced by a common cause or unmeasured confounder, SEM #3.

Each of these structural hypotheses can be independently evaluated for how well they represent the underlying data, and the best-fit solution can be admitted for use.

Table 2 summarizes the goodness-of-fit test results for those three models, obtained by comparing the implied covariance matrix (derived from structural equation models) with the empirically observed data covariance[10]. A higher p-value suggests a good fit of the data given the model (both SEM and graph structure), while a lower p-value may suggest inconsistency between the data and model. As shown, model SEM #2 has the highest p-value, comparing with two other models. Since the only difference among the three models is the configuration of the arrow between *Support* and *Affect*, this result suggests that the data provides more evidence of *Affect* as an antecedent rather than as a consequence of *Support*. In other words, parents seem to use their observation of child's affective experience as input to decide on the types of support they provide.

Table 3 summarizes the estimated parameters and related statistical testing results for the three structural equation models SEM #1, SEM #2 and SEM #3 corresponding to the graph structures in Figure 6 subplots (b), (c) and (d), respectively. All estimated coefficients are positive and significantly different from zero as evidenced by very low p-values.[11] Focusing on

[10]To be precise, the test statistics, due to (Bollen, 1989), is derived from minimizing a function of maximum-likelihood (FML). When FML is minimized, the distance from the minimum point to zero is proportional to the χ^2 of the model. Latent variables are allowed as long as the model is linear Gaussian.

[11]This p-value is derived for the null-hypothesis that the edge coefficient is zero. Unlike the model p-value discussed above, lower p-values here suggest greater utility of the particular edge, as it supports rejecting the null-hypothesis of zero value of edge coefficient.

Table 3: Estimated parameters from three alternative structural equation models corresponding to graph structures in Figure 6(b), (c), and (d).

Model	From	To	Type	Value	Std.Err	T-Value	P-value
SEM #1	Profile	Affect	Edge Coef.	0.51	0.1298	3.9376	0.0002
	Support	Affect	Edge Coef.	0.32	0.1121	2.8931	0.0056
	Tot	Support	Covariance	0.62			
SEM #2	Profile	Affect	Edge Coef.	0.71	0.1364	5.1764	≤ 0.0001
	Affect	Support	Edge Coef.	0.44	0.1280	3.4102	0.0013
	Tot	Support	Covariance	0.49			
SEM #3	Profile	Affect	Edge Coef.	0.58	0.1364	4.2641	0.0001
	Tot	Support	Covariance	0.54			
	Affect	Support	Covariance	0.48			

SEM #2 model which appears to be the best-fit model among the three, we note that the sign between *Affect* and *Support* is positive, which suggests that parents' observation of a higher level of child's struggle may lead them to activate more intensive types of support. Often, this means the cognitive support that is explicit and elaborate which may take more time to implement than less involved forms of support. It is also interesting to note that there is a positive coefficient for the path from *Profile* to *Affect*. This suggests that a child with a high profile score (i.e., higher Grit-scale score, low self-efficacy, and with more introvert personality) may experience more struggles, which in turn is likely to trigger higher level of support from their parent.

ANALYSIS OF THE 7-FACTOR MODEL This analysis expands the input space of our causal reasoning by including three additional composite variables:

- *Interact*: this composite variable describes parent's talk relative to the child's talk, which is an indirect measure of the amount of support since most of parent's talk involves coaching. This variable provides complementary information to the existing composite variable *Support* included in the 4-factor model, which primarily captures the composition of a mix of the types of support;
- *Stress*: this composite variable is comprised of variables derived from the moment-to-moment dynamics of the reserve level, providing complementary information to the session level marginal distribution of CD and EC states as reflected in the *Affect* variable;
- *Assess*: this composite variable aims to encode parents' assessment of child participants' cognitive-affective experience during and toward the end of a session, using session logs completed at the time of recording.

Subplot (a) in Figure 7 depicts the equivalence class of graph discovered by the same search algorithm (GFCI) and parameters as in the 4-factor model. From this graph, we observed several high-level patterns of connectivity with the introduction of new variables. For example, the discovered model reveals that *Affect* is a direct cause of *Stress* without any latent confounders (as indicated by the solid directional arrow). The relationship between *Interact* and *Support* however carries more uncertainty: *Interact* could be a cause of *Support*, but it is also possible that those two variables are related through a common confounder. In addition, we note that *Affect*

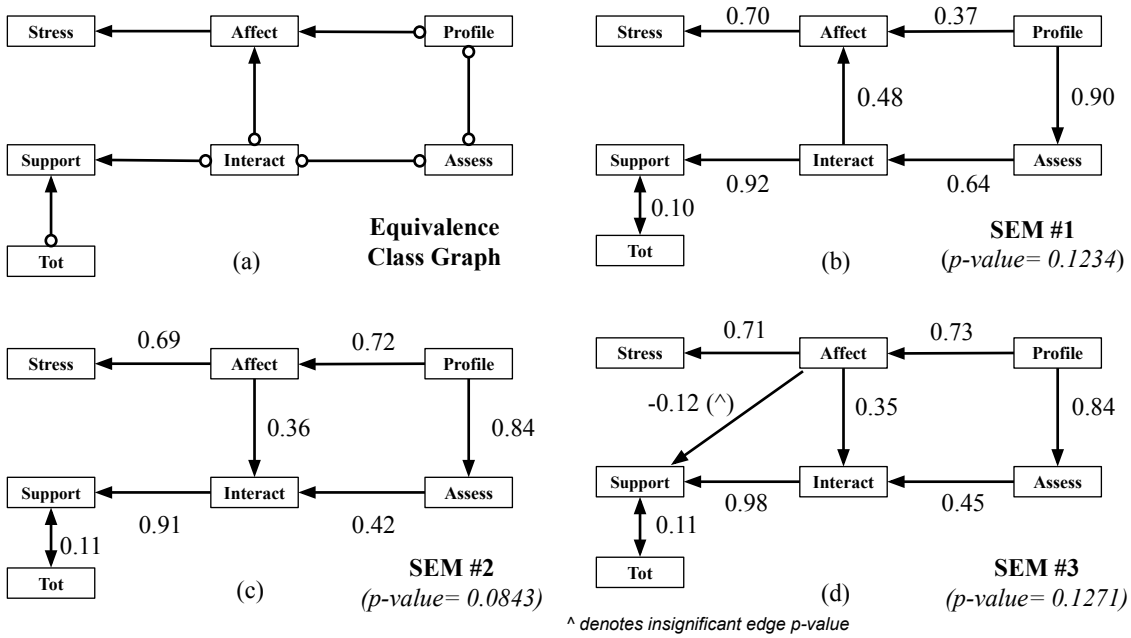

Figure 7: Causal models for the 7-factor model. Subplot(a) is the equivalence class graph discovered from the search algorithm in TETRAD; subplot (b),(c), and (d) are alternative refined models taking into account domain knowledge and hypothesis to test. The numeric values on the edge are estimated edge coefficients from the structural equation models. All edge-coefficients are significantly different from zero except for the one noted.

variable interacts directly with *Interact* variable but not with *Support* given the missing edge between the two. Likewise, *Assess* does not directly interact with *Affect*; it, however, interacts through *Profile*. Meanwhile, *Assess* is also related to the *Interact* variable, but the directionality cannot be determined without further hypothesis testing, in addition to the possibility of involving a common confounder.

We then derived alternative causal models which are largely consistent with the patterns in Figure 7(a). We incorporated the following specifications as informed by our understanding of the data: (1) a bi-directional arrow between *Support* and *Tot* as we believe there could be an unmeasured confounder reflective of the problem level of difficulty level; (2) a directional arrow out of *Profile* as it is believed to be an exogenous variable. Given those imposed constraints, we are still interested in testing the three remaining alternative hypotheses regarding the relationship between *Affect* and support-related variables (*Support* and *Interact*), which are reflected in the configuration of corresponding edges or arrows in subplots shown in Figure 7(b)-(d).

Since the discovered causal graphs in Figure 7(a) does not support the direct relationship between *Affect* and *Support*, as it did in the 4-factor model, we test the relationship between *Affect* and *Interact* instead. Since the result from the 4-factor model supports evidence of the causal linkage from *Affect* to *Support*, we test an additional hypothesis by adding a direct edge from *Affect* to *Support*. The hypotheses to be tested and the corresponding graphs are:

- H1 as in subplot 7(b): *Interact* influences *Affect*, yielding structural equivalence model SEM #1;

- H2 as in subplot 7(c): *Affect* influences *Interact*, SEM #2;
- H3 as in subplot 7(d): *Affect* influences both *Interact* and *Support*, SEM #3.

Table 4 summarizes the goodness-of-fit test for the three alternative SEM models, and Table 5 summarizes the estimated parameters and related statistical testing results for the three structural equation models.

We observe that we cannot reject either of the alternative formulations of directional influence between *Affect* and *Interact* (Figure 7(b) and (c)), as indicated by the insignificant p-values reported for them in Table 4. This suggests that those models are able to explain data reasonably well. By examining the signs of the edge coefficients and their statistical significance shown in Table 5, we note that SEM #1 model that assumes a direct positive causal pathway from *Interact* (intervention) to *Affect*, suggesting that more parent's talking may in fact give rise to more struggles for the child. This seems to reflect sub-optimal scenarios where parents are rendering excessive help. We note this kind of behavior often occurred in sessions where parents took early control in the problem-solving process, leaving little room for the child to be engaged independently thus reducing their likelihood of experiencing positive affect such as Engaged Concentration. In some extreme cases, intensive support not only fails to reduce child's struggles, but it may also backfire, causing more frustration for the child when they find the parents' help disruptive rather than helpful.

Similarly, from model fitting results for SEM #2 and #3, we note that *Affect* may also be the cause of parents' support (both in terms of the extent and type), in other words, parents may take into account child's affect when deciding to provide support. This finding is generally consistent with the results from the 4-factor model; it is, however, worth noticing a slight difference. While the 4-factor model identifies *Affect* as the direct cause to the types of support (*Support*), here the model seems to favor the pathway from *Affect* to the amount of support (*Interact*) rather than to type of support. In fact, as shown in SEM #3, the additional edge from *Affect* to *Support* seems to be a weak one with insignificant (large) p-value as shown in Table 5, even though it improves the model fit slightly. This finding suggests that the amount of support is a more actionable variable from parents' point of view. Instead of deciding on the nuances of the type of support, parents' simply need to decide whether or not he or she likes to intervene by engaging in or withholding from talking, which seems to be aligned with human decision making intuitions. As such, we speculate that the pathway from *Affect* to *Support* in the 4-factor model indeed reflects the indirect pathway from *Affect* to *Interact*, and then onto *Support*, as explicitly demonstrated in the 7-factor model.

Compared with the 4-factor alternative, the 7-factor model provides additional insights: (1) Complementary composite variables such as *Affect* and *Stress* are found to have definite positive

Table 4: Goodness-of-fit testing statistics for three alternative structural equation models corresponding to graph structures in Figure 7(b), (c), and (d).

Model	SEM #1	SEM #2	SEM #3
Degrees of Freedom	14	14	13
χ^2	20.2181	21.7320	18.8729
P Value	0.1234	0.0843	0.1271
BIC Score	-35.3660	-33.8521	-32.7409

Table 5: Estimated parameters from the structural equation model corresponding to graph structures in Figure 7(b), (c), and (d).

Model	From	To	Type	Value	SE	T-value	P-value
SEM #1	Affect	Stress	Edge Coef.	0.6960	0.073	9.6	≤ 0.0001
	Profile	Affect	Edge Coef.	0.3660	0.133	2.747	0.0080
	Interact	Support	Edge Coef.	0.9530	0.069	13.765	≤ 0.0001
	Assess	Interact	Edge Coef.	0.6410	0.114	5.612	≤ 0.0001
	Interact	Affect	Edge Coef.	0.4840	0.124	3.891	≤ 0.0001
	Profile	Assess	Edge Coef.	0.9070	0.114	7.941	≤ 0.0001
	Tot	Support	Covariance	0.1020			
SEM #2	Profile	Assess	Edge Coef.	0.8982	0.1143	7.8602	≤ 0.0001
	Assess	Interact	Edge Coef.	0.4239	0.1224	3.463	0.0011
	Affect	Interact	Edge Coef.	0.3614	0.1112	3.2502	0.0020
	Profile	Affect	Edge Coef.	0.7160	0.1364	5.2483	≤ 0.0001
	Affect	Stress	Edge Coef.	0.6883	0.0725	9.4889	≤ 0.0001
	Interact	Support	Edge Coef.	0.9085	0.0692	13.1291	≤ 0.0001
	Tot	Support	Covariance	0.1060			
SEM #3	Profile	Affect	Edge Coef.	0.7321	0.1364	5.3665	≤ 0.0001
	Affect	Interact	Edge Coef.	0.3513	0.1112	3.1598	0.0026
	Assess	Interact	Edge Coef.	0.4491	0.1224	3.669	0.0006
	Affect	Support	Edge Coef.	-0.1214	0.0772	-1.5718	0.1221
	Profile	Assess	Edge Coef.	0.8402	0.1143	7.3533	≤ 0.0001
	Interact	Support	Edge Coef.	0.9790	0.0829	11.8162	≤ 0.0001
	Affect	Stress	Edge Coef.	0.7115	0.0725	9.809	≤ 0.0001
	Tot	Support	Covariance	0.1124			

directional causal relationships. We note that a higher prevalence of cognitive disequilibrium episodes, captured in *Affect*, will lead to a high value of stress index (e.g., the higher likelihood of reaching a lower point of the valley of the emotional roller-coaster); (2) We note a positive relationship between the pair of variables describing parent's support: *Interact* (the amount and proportion of parent's talk) has a direct positive influence on the type of *Support*. In other words, more parent talk leads to a higher likelihood that they are rendering cognitive support. This is consistent with our definition of support types: cognitive support is more elaborate and thus more time consuming than meta-cognitive or social/emotional types.

We also note an interesting role the composite variable *Profile* plays in this model. Similar to the 4-factor model, it has influence on the *Affect* variable. Interestingly, it also contributes to the intervention variable *Interact* indirectly via *Assess* variable. It is possible that parents may take into account child's profile (e.g., personality, self-efficacy, etc.) in assessing the child's affective experience and this assessment is then reflected in the *Assess* variable. Furthermore, this assessment may in turn determine their intervention strategies, which is reflected in the amount of talking they contribute to the coaching sessions. This causal pathway seems to suggest that it is not the child's profile itself, but indeed parent interpretation of the child's profile (which is reflected in their subjective assessment of the child's affective experience) that influences their support decisions.

7. DISCUSSION

We have explored an analysis pipeline that identifies patterns of correlations between functionally distinct feature sets of multimodal behavioral data and uses the derived composite variables as factors in causal analysis. The proposed methodology combines sparse multiple canonical correlation analysis and causal structure discovery and inference methods. This pipeline is suitable for analyzing datasets comprising a relatively small number of samples compared to the dimensionality of the feature space. In particular, it caters to representing variables that can be grouped into multiple categories from which we are interested in discovering and validating causal relationships. This framework can facilitate the interpretation of interaction processes readily discovered from empirical data, making it relatively easier to incorporate domain knowledge or intuitive beliefs. This type of data is often seen in multimodal human-human interaction studies, where data collected asynchronously through different sensing modalities, and at various temporal resolutions.

7.1. SUMMARY OF FINDINGS

We demonstrated our methodology using a multimodal one-on-one coaching dataset. There are several insights we were able to derive from this analysis.

Firstly, we note a clear causal pathway between the group of variables describing parents' support and another group representing the child's cognitive-affective experience. It is worth pointing out that these two groups of variables are derived from two different sensing modalities and are independently annotated. Specifically, affect related descriptors are mainly informed by visual channels (e.g., facial expressions and gross body movements), and support-associated descriptors are derived from audio streams. Based on the collective evidence from multiple model-fitting exercises, what comes to light is the insight that supports a two-way interaction between those two streams of data. Also, we note that the richer 7-factor model identifies *Interact* variable (a proxy for the amount of support) as a more actionable variable than the type of support, which was not captured by the less potent 4-factor model. This distinction demonstrates the benefit of including more variables in causal modeling, given a sufficient supply of reliable and relevant empirical data.

Secondly, we note the causal pathway from *Profile* to *Affect* and, indirectly, to *Support*. Especially, as evidenced in SEM #1 variant of the 7-factor model, a child's affective-cognitive experience is a function of both the stable factors (e.g., a child's profile) and the situational factors (e.g., parent's intervention that may be adaptive to the given coaching session). This empirical finding informs us about how parents approach supporting their child during one-on-one coaching, and highlights the kind of information they might have took into account in their support decisions.

It should be noted that the causal models discussed here are dependent on the annotation labels for child's affective/cognitive states and parents' support types. As such, the uncertainty arising from the annotations may influence the resulting model structures, parameters and interpretation.

7.2. IMPLICATIONS

This research proposed a multimodal analytics framework that allows researchers to gain insights into the "black box" of the complex, multi-party, fine-grained multimodal interactions

data streams collected from human-to-human one-on-one coaching. The unification of a sparse modelling procedure and causal analysis framework brings transparency into this complex process. In addition to a concrete understanding of “what" is happening, we now have the tool to ask interesting “why" questions. This additional benefit in transparency is essential to make multimodal learning analytics accessible to multiple stakeholders such as teachers, students and researchers. When used appropriately and creatively, it also has the potential to open up opportunities to turn the learned causal knowledge into actionable insights to ultimately improve learning outcomes.

In addition, this research contributes to several research communities. Firstly, it enriches the research portfolios in mathematics education, particularly in the areas of affect dynamics, self-regulation, and effective intervention with young children. Researchers in this community may use the methodology framework introduced in this paper to derive deep insights into audio and video data typically collected, ultimately to guide the design of technology-enhanced interventions to improve learning outcomes. Secondly, this research contributes to the affective computing community by exploring analytical methods operating at the reasoning layer on top of the perceptual layers, which has been the focus of most of the recent work in this space. Thirdly, this research contributes to the relatively sparse literature in multimodal learning analytics by providing a balanced view of the teacher and student interactions with a data set collected in naturalistic home environments.

7.3. Limitations and Future Work

There are several limitations to this work that should motivate further investigations. Firstly, the size of the dataset used is relatively small, and the subject pool is not overly diverse, limiting our ability to explore culture or ethics-related factors in the model reliably. As part of the future work, we intend to expand the study to recruit additional subjects from more diverse backgrounds. Secondly, the study collected only audio data instead of multimodal audio/video streams from parents. While this decision reflects the compromise with the logistic complexity of data collection, it limits the opportunity to account for multimodal bi-directional interactions between children and parents. Thirdly, instead of using experienced tutors as most traditional tutoring studies do, we let parents take up the coaching roles. While parents possibly excel at interpreting the child’s affective experience, they may not be the ideal tutors capable of rendering optimal supports. However, it does not hinder us from applying the methods discussed here to future data collected with the participation of experienced tutors. Fourthly, this work analyzes the causal relationships at the level of individual coaching session. While this is a first step towards the complete understanding of the intertwining processes of a child’s cognitive-affective experience and parent’s coaching decisions, further work is necessary to model the moment-by-moment causal relationships by exploiting temporal relationships of various events and variables at a granular level.

8. Conclusion

With a multimodal one-on-one coaching dataset, we introduced a methodological framework to support causal inference and discovery of causal relationships among groups of variables that describe parents’ coaching decisions, children’s cognitive-affective experiences, and children’s individual stable factors. When complemented with computational models for recog-

nizing learning and teaching-related constructs at the *perceptual layer*, we envision the causal chains based on *reasoning layer* analysis may be scaled up to large datasets. This will augment our ability to uncover the complex relationships between teachers' actions and students' responses. We believe that causal understanding is crucial in achieving the ambitious goal of designing a brilliant teaching machine with human-like reasoning that could provide personalized and adaptive supports to optimize learning outcomes for all.

9. ACKNOWLEDGMENTS

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. In addition, the authors would like to thank Mononito Goswami, Qianou Ma and Eva Gjekmarkaj for their talented and dedicated research assistance.

REFERENCES

AMBADY, N. AND ROSENTHAL, R. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin 111,* 2, 256–274.

BAKER, R. S., D'MELLO, S. K., RODRIGO, M. M. T., AND GRAESSER, A. C. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies 68,* 4, 223–241.

BAKER, R. S. AND YACEF, K. 2009. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining 1,* 1, 3–17.

BANDURA, A. 2006. Guide for constructing self-efficacy scales. In *Self-efficacy Beliefs of Adolescents*, F. Pajares and T. Urdan, Eds. Age Information Publishing, Greenwich, 307–337.

BOLLEN, K. A. 1989. *Structural Equations with Latent Variables*. Wiley, New York.

CHEN, L., LI, X., XIA, Z., SONG, Z., MORENCY, L.-P., AND DUBRAWSKI, A. 2016. Riding an emotional roller-coaster: A multimodal study of young child's math problem solving activities. In *Proceedings of the 9th International Conference on Educational Data Mining*, T. Barnes, M. Chi, and M. Feng, Eds. 38–45.

CSIKSZENTMIHALYI, M. 2013. *Flow: The Psychology of Happiness*. Random House.

CUKUROVA, M., GIANNAKOS, M., AND MARTINEZ-MALDONADO, R. 2020. The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology 51,* 5, 1441–1449.

DE CARVALHO, W. F., COUTO, B. R. G. M., LADEIRA, A. P., GOMES, O. V., AND ZARATE, L. E. 2018. Applying causal inference in educational data mining: A pilot study. In *Proceedings of the 10th International Conference on Computer Supported Education*, B. M. McLaren, R. Reilly, S. Zvacek, and J. O. Uhomoibhi, Eds. 454–460.

DI LEO, I. AND MUIS, K. R. 2020. Confused, now what? A cognitive-emotional strategy training (CEST) intervention for elementary students during mathematics problem solving. *Contemporary Educational Psychology 62*, 101879.

DI LEO, I., MUIS, K. R., SINGH, C. A., AND PSARADELLIS, C. 2019. Curiosity... confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology 58*, 121–137.

D'MELLO, S., DALE, R., AND GRAESSER, A. 2012. Disequilibrium in the mind, disharmony in the body. *Cognition & Emotion 26,* 2, 362–374.

D'MELLO, S. K. AND KORY, J. 2015. A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys 47,* 3, 1–36.

DRACHSLER, H. AND SCHNEIDER, J. 2018. JCAL special issue on multimodal learning analytics. *Journal of Computer Assisted Learning 34,* 4, 335–337.

DU BOULAY, B. AND LUCKIN, R. 2016. Modelling human teaching tactics and strategies for tutoring systems: 14 years on. *International Journal of Artificial Intelligence in Education 26,* 1, 393–404.

DUCKWORTH, A. L. AND QUINN, P. D. 2009. Development and validation of the short grit scale (GRIT–S). *Journal of Personality Assessment 91,* 2, 166–174.

DWECK, C. S. 2008. *Mindset: The New Psychology of Success*. Random House Digital, Inc.

D'MELLO, S. AND GRAESSER, A. 2012. Dynamics of affective states during complex learning. *Learning and Instruction 22,* 2, 145–157.

ELLIOT, A. J. AND MURAYAMA, K. 2008. On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology 100,* 3, 613–628.

EMERSON, A., CLOUDE, E. B., AZEVEDO, R., AND LESTER, J. 2020. Multimodal learning analytics for game-based learning. *British Journal of Educational Technology 51,* 5, 1505–1526.

FANCSALI, S. 2014. Causal discovery with models: Behavior, affect, and learning in cognitive tutor algebra. In *Proceedings of the 7th International Conference on Educational Data Mining*, J. C. Stamper, Z. A. Pardos, M. Mavrikis, and B. M. McLaren, Eds. 28–35.

FANCSALI, S. 2015. Confounding carelessness? Exploring causal relationships between carelessness, affect, behavior, and learning in cognitive tutor algebra using graphical causal models. In *Proceedings of the 8th International Conference on Educational Data Mining*, O. C. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, M. C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, and M. C. Desmarais, Eds. 508–511.

GIANNAKOS, M. N., SHARMA, K., PAPPAS, I. O., KOSTAKOS, V., AND VELLOSO, E. 2019. Multimodal data as a means to understand the learning experience. *International Journal of Information Management 48*, 108–119.

GOSWAMI, M., CHEN, L., AND DUBRAWSKI, A. 2020. Discriminating cognitive disequilibrium and flow in problem solving: A semi-supervised approach using involuntary dynamic behavioral signals. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 420–427.

GRAESSER, A. C., PERSON, N., HARTER, D., AND TUTORING RESEARCH GROUP. 2001. Teaching tactics in AutoTutor. *International Journal of Artificial Intelligence in Education 12*, 257–279.

GRAWEMEYER, B., MAVRIKIS, M., HOLMES, W., GUTIÉRREZ-SANTOS, S., WIEDMANN, M., AND RUMMEL, N. 2017. Affective learning: improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction 27,* 1, 119–158.

HUTT, S., KRASICH, K., MILLS, C., BOSCH, N., WHITE, S., BROCKMOLE, J. R., AND D'MELLO, S. K. 2019. Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction 29,* 4, 821–867.

JOHN, O. P. AND SRIVASTAVA, S. 1999. The big five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and research*, O. P. John and R. W. Robins, Eds. Guilford, 102–138.

KAPUR, M. 2014. Productive failure in learning math. *Cognitive Science 38,* 5, 1008–1022.

KOEDINGER, K. R. AND ALEVEN, V. 2007. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review 19,* 3, 239–264.

KOEDINGER, K. R., MCLAUGHLIN, E. A., JIA, J. Z., AND BIER, N. L. 2016. Is the doer effect a causal relationship? How can we tell and why it's important. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, D. Gasevic, G. Lynch, S. Dawson, H. Drachsler, and C. P. Rosé, Eds. 388–397.

LEHMAN, B., D'MELLO, S., CADE, W., AND PERSON, N. 2012. How do they do it? Investigating dialogue moves within dialogue modes in expert human tutoring. In *International Conference on Intelligent Tutoring Systems*, S. A. Cerri, W. J. Clancey, G. Papadourakis, and K. Panourgia, Eds. Springer, 557–562.

LINNENBRINK-GARCIA, L., DURIK, A. M., CONLEY, A. M., BARRON, K. E., TAUER, J. M., KARABENICK, S. A., AND HARACKIEWICZ, J. M. 2010. Measuring situational interest in academic domains. *Educational and Psychological Measurement 70,* 4, 647–671.

MARTINEZ-MALDONADO, R., KAY, J., BUCKINGHAM SHUM, S., AND YACEF, K. 2019. Collocated collaboration analytics: Principles and dilemmas for mining multimodal interaction data. *Human–Computer Interaction 34,* 1, 1–50.

MILLER, I. W. AND NORMAN, W. H. 1979. Learned helplessness in humans: A review and attribution-theory model. *Psychological Bulletin 86,* 1, 93.

MU, S., CUI, M., AND HUANG, X. 2020. Multimodal data fusion in learning analytics: A systematic review. *Sensors 20,* 23, 6856.

OCHOA, X. 2017. Multimodal Learning Analytics. In *The Handbook of Learning Analytics*, 1 ed., C. Lang, G. Siemens, A. F. Wise, and D. Gasevic, Eds. Society for Learning Analytics Research, Alberta, Canada, 129–141.

PEKRUN, R. AND STEPHENS, E. J. 2010. Achievement emotions: A control-value approach. *Social and Personality Psychology Compass 4,* 4, 238–255.

PINTRICH, P., SMITH, D., GARCÍA, T., AND MCKEACHIE, W. 1991. A manual for the use of the motivated strategies for learning questionnaire (MSLQ). Tech. rep., University of Michigan, Ann Arbor, MI.

PORAYSKA-POMSTA, K., MAVRIKIS, M., AND PAIN, H. 2008. Diagnosing and acting on student affect: The tutor's perspective. *User Modeling and User-Adapted Interaction 18,* 1-2, 125–173.

PRIETO, L. P., SHARMA, K., DILLENBOURG, P., AND JESÚS, M. 2016. Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, D. Gasevic, G. Lynch, S. Dawson, H. Drachsler, and C. P. Rosé, Eds. 148–157.

PRIETO, L. P., SHARMA, K., KIDZINSKI, Ł., RODRÍGUEZ-TRIANA, M. J., AND DILLENBOURG, P. 2018. Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning 34,* 2, 193–203.

SANTOS, O. C., SANEIRO, M., SALMERON-MAJADAS, S., AND BOTICARIO, J. G. 2014. A methodological approach to eliciting affective educational recommendations. In *2014 IEEE 14th International Conference on Advanced Learning Technologies*. IEEE, 529–533.

SCHEINES, R., SILVER, E., AND GOLDIN, I. M. 2014. Discovering prerequisite relationships among knowledge components. In *Proceedings of the 7th International Conference on Educational Data Mining*, J. C. Stamper, Z. A. Pardos, M. Mavrikis, and B. M. McLaren, Eds. 355–356.

SCHOENFELD, A. H. 2016. Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (reprint). *Journal of Education 196,* 2, 1–38.

SHARMA, K. AND GIANNAKOS, M. 2020. Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology 51,* 5, 1450–1484.

TSUKAYAMA, E., DUCKWORTH, A. L., AND KIM, B. 2013. Domain-specific impulsivity in school-age children. *Developmental Science 16,* 6, 879–893.

WITTEN, D. M. AND TIBSHIRANI, R. J. 2009. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical Applications in Genetics and Molecular Biology 8,* 1, 1–27.

WORSLEY, M. AND BLIKSTEIN, P. 2018. A multimodal analysis of making. *International Journal of Artificial Intelligence in Education 28,* 3, 385–419.

YADEGARIDEHKORDI, E., NOOR, N. F. B. M., AYUB, M. N. B., AFFAL, H. B., AND HUSSIN, N. B. 2019. Affective computing in education: A systematic review and future research. *Computers & Education 142*, 103649.

ZHU, G., XING, W., COSTA, S., SCARDAMALIA, M., AND PEI, B. 2019. Exploring emotional and cognitive dynamics of knowledge building in grades 1 and 2. *User Modeling and User-Adapted Interaction 29,* 4, 789–820.

ZIMMERMAN, B. J. 2000. Attaining self-regulation: A social cognitive perspective. In *Handbook of Self-Regulation*, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds. Academic Press, San Diego, 13–39.

APPENDIX

A. SESSION LOG COMPLETED BY PARENTS

A.1. ACTIVITY LOG

1. What was the activity your child was doing immediately before working on the given math problem? (e.g., eating dinner, other math homework, playing soccer, etc.)

2. Approximately for how long was your child engaged in that activity?

Due to the sparsity of the data, this part of the log is not coded into the session-level descriptors

A.2. PARENT'S ASSESSMENT OF CHILD'S EMOTIONAL EXPERIENCE DURING AND AFTER EACH SESSION

1. During this session, how frustrated do you believe your child became? (1=not frustrated; 5= very frustrated)

2. During this session, how much do you think your child was enjoying working on the problem? (1=not enjoyed 5= very enjoyed)

3. During the session, how engaged do you think your child was while working on the problem? (1= disengaged; 5= very engaged)

4. At the end of the session, how would you describe how your child felt? Check all that apply.

    - Accomplished
    - Joyful
    - Frustrated
    - Confused
    - Surprised
    - Neural
    - Other (please specify)

B. List of Subject-level Descriptors

Table 6: Subject-level descriptors.

Name	Description
A	Achievement score
E	Effort Regulation score
G	Grit scale score
H	Help-seeking score
M	Math Interest score
PA	Personality - Agreeableness sub-score
PC	Personality - Conscientiousness sub-score
PE	Personality - Extroversion sub-score
PN	Personality - Neuroticism sub-score
PO	Personality - Openness sub-score
S_C	Self-control score - Child assessment
S_P	Self-control score - Parent assessment
SE	Self-efficacy score

C. List of Session-level Descriptors

Table 7: Session-level descriptors.

Category	Name	Description
Time	Tot	Time on Task; in minutes
Affect	n_CD	Number of segments labeled as CD
Affect	n_EC	Number of segments labeled as EC
Affect	CD_EC_ratio	The ratio between CD segments and EC segments
Affect	Pctg_CD	Percentage of CD segments
Affect	Pctg_EC	Percentage of EC segments
Affect	Pctg_neutral	Percentage of Neutral segments
Affect	Pctg_CD2	Percentage of CD and Neutral segments combined
Affect	Pctg_EC2	Percentage of EC and Neutral segments combined
Support	C_count	Number of cognitive support utterances
Support	C_length	Cumulative duration of cognitive support utterances
Support	M_count	Number of meta-cognitive support utterances
Support	M_length	Cumulative duration of meta-cognitive support utterances
Support	S_plus_count	Number of positive social/emotional support utterances
Support	S_minus_count	Number of negative social/emotional support utterances
Support	M_prop	Proportion of meta-cognitive support in terms of counts
Support	M_prop_duration	Proportion of meta-cognitive support in terms of duration
Support	C_prop_tot	Cumulative cognitive support; normalized by Tot
Support	M_prop_tot	Cumulative meta-cognitive support; normalized by Tot
Support	CM_count	The number of cognitive and meta-cognitive support

Continuation of Table 7

Category	Name	Description
Stress	Global_low	Global low of reserve(this could be a negative quantity)
Stress	Global_high	Global high of reserve
Stress	Global_delta	Global high minus global low
Stress	Max_up	The longest non-decreasing streaks
Stress	Max_down	The longest non-increasing streaks
Stress	Num_up	Number of non-decreasing streaks lasting 60+ secs
Stress	Num_down	Number of non-increasing streaks lasting 60+ secs
Stress	Max_CD_length	The longest CD streaks
Stress	Max_EC_length	The longest EC streaks
Stress	Num_CD	Number of CD streaks that last at least 30 seconds
Stress	Num_EC	Number of EC streaks that last at least 30 seconds
Interact	Child_eyeGaze_count	Number of child's eye gazes toward parent
Interact	Child_eyeGaze_duration	Cumulative duration of child's eye gaze toward parent
Interact	Child_talk_count	Number of child's utterance
Interact	Child_talk_duration	Cumulative duration of child's utterance
Interact	Parent_talk_count	Number of parent's utterance
Interact	Parent_talk_duration	Cumulative duration of parent's utterance
Interact	Parent_talk_pctg	Duration wise,proportion of talk that is from parent
Assess	During_frustrated	Parent's rating of child's frustration during session
Assess	During_enjoy	Parent's rating of child's enjoyment during session
Assess	During_engaged	Parent's rating of child' engagement during session
Assess	End_accomplish	Parent's rating of child's accomplished feeling end of session
Assess	End_joy	Parent's rating of child's joy end of session
Assess	End_frustrated	Parent's rating of child's frustration end of session
Assess	End_confused	Parent's rating of child's confusion end of session
Assess	End_surprised	Parent's rating of child's surprise end of session
Assess	End_neutral	Parent's rating of child's neutral feeling end of session

C.1. CODING GUIDES FOR CHILD'S COGNITIVE/AFFECTIVE STATES

CD=Cognitive Disequilibrium ; EC=Engagement Concentration

Contextual Information		**Off Task**	**Disengaged**	**Engaged**		
Main Patterns	**Secondary Patterns**			**CD**	**Neutral**	**EC**
Active (child in control)	Child exhibits overt productive behaviors (e.g., think-aloud or writing)	NA	Unlikely	Verbal Cues [Note A]	No obvious clues suggestive of CD	Coherent, cohesive or fast talking speed; fast writing speed
	Child is thinking in silence (no talking or writing)	NA	Unlikely	Visual Cues [Note B]	No obvious clues suggestive of CD	Unlikely
Passive (parent in control)	Mainly parent monologue with child passively listening	NA	Child shows signs of distraction (e.g., look around) or boredom (e.g., yawning)	Visual cues [Note B] or Verbal cues [Note A]	No obvious clues suggestive of CD	Unlikely
Interactive (equal control)	Parent and child engaged in dialogue	NA	Unlikely	Visual cues [Note B] or Verbal cues [Note A]	Most likely	Less likely
Off Task	Irrelevant with current tasks (e.g. device malfunction or lookup an answer)	Yes	NA	NA	NA	NA

Notes:

(A): Examples of verbal cues: "I don't understand"; "that does not make sense";
(B): Examples of visual cues: Confused or frustrated facial expression; scratching head or biting pen

C.2. CODING GUIDES FOR PARENTS' SUPPORT TYPES

Code Level1	Code Level2	Description	Definition	Examples
F	F+	Positive Feedback	Positive Feedback	“Yes"
F	F-	Negative Feedback	Negative Feedback	“No"
F	F	Neutral Feedback	Back channel response without clear indication of positive or negative	“Okay";“Uhm"
C	C	Cognitive Support	Direct and targeted support such as asking questions specific to the given problem	Re-read questions;Make corrections; Explain/clarify; Scaffolding; Ask questions (Infrequently, C- for detailed but unhelpful or misleading support)
M	M	Meta-cognitive Support	Indirect support, hints on general problem solving strategy, without specific or detailed hints	High level scaffolding, may refer to general problem solving strategies (e.g. re-read question, draw a diagram, etc.)
S	S+	Positive Social or Emotional Support	Praise or reassurance/encouragement	“You are doing great!"
S	S-	Negative Social or Emotional support	Utterance that likely to discourage child and generate negative feelings	“How does this make any sense?”, “This should not be a very hard problem.”
O	O	Others	Irrelevant/Off Task Utterance	Misc.
H	H	Help Dynamics	Utterance related to intentions such as offering help, refusing to help, initiation of help etc.	“Do you need help?” “You can do it by yourself”(Note A)

Note: Utterances may look similar to C or M, however, since they appear at the beginning of the assistance phase, we annotate them as H instead of C or M to identify their unique functions in the interaction.