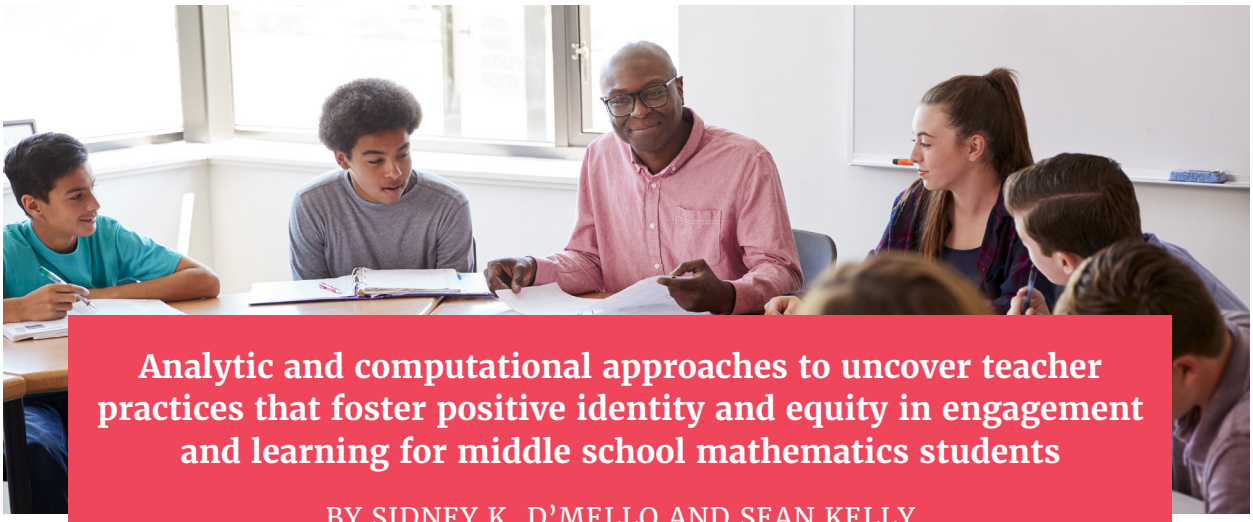

Analytic and computational approaches to uncover teacher practices that foster positive identity and equity in engagement and learning for middle school mathematics students

BY SIDNEY K. D'MELLO AND SEAN KELLY

RESEARCH SNAPSHOT | OCTOBER 2021

From carefully crafted messages to flippant remarks, warm expressions to unfriendly tones, teachers' behaviors set the tone, expectations, and attitudes of the classroom.

Though not always intentional, certain teacher behaviors risk marginalizing students with stigmatized identities, especially with respect to gender, race and ethnicity, and socioeconomic status. Other teachers create a normative environment of inclusion and widespread engagement through effective discourse practices, among other strategies. Because measurement can be one catalyst of change, we designed this study to identify the ways in which teachers foster motivation, positive identity, and a strong sense of belonging through inclusive messaging and other nonverbal interactions.

RESEARCH TEAM

- Sidney K. D'Mello (PI), University of Colorado Boulder
- Sean Kelly (Co-I), University of Pittsburgh
- Stephanie Wormington (Co-I), Center for Creative Leadership
- Hadassah Muthoka, University of Colorado Boulder
- Nicholas Hunkins, University of Colorado Boulder
- Erin Vines, University of Colorado Boulder
- Julianna Harris, University of Colorado Boulder
- Caroline O'Reilly, University of Colorado Boulder
- Amanda Michaels, University of Colorado Boulder

Areas of expertise: Educational psychology, educational policy, sociology of education, and computer science

Key Findings

- We identified a set of prevalent teacher discourse practices likely to affect students' psychological and academic outcomes. These practices varied substantially from lesson to lesson and from teacher to teacher.
- We trained observers to achieve high levels of reliability in identifying these discourse dimensions from classroom video and investigated the feasibility of automatic discourse coding to scale up similar analyses.
- We established the validity of our new fine-grained measures in comparison to existing teacher observation protocols and with student measures of belonging and achievement.

Using video recordings of 6th to 8th grade mathematics classes, student self-report questionnaires, existing evaluations of teacher practice, and achievement data from the Measures of Effective Teaching (MET) project, we aim to identify teacher verbal and nonverbal behaviors that are related to students' psychological and academic outcomes, particularly for students with stigmatized identities.

Study Design

Our study is observational and involves a novel coding of videos. We strategically sampled videos to code such that there was variability in students' perceptions of the

student experience RESEARCH NETWORK

Student Experience Research Network (SERN), formerly known as Mindset Scholars Network, envisions an education system in which every student experiences respect as a valued person and thinker, so that they can learn and thrive. To this end, SERN connects people and ideas across research, practice, and policy to advance scientific knowledge and inform decision-making by education system and institution leaders.

Sample

Our sample consists of 6th to 8th grade mathematics students and their teachers from the extant Measures of Effective Teaching (MET) data set. The sample is derived from 156 recorded classroom videos in 73 class sections. Our sampling is stratified by student perception scores (from TRIPOD assessment), oversampling high and low rated class sections. The analyzed sample includes data from approximately 1,400 students, of which 52% are male, 47% are eligible for free- or reduced-price lunch, 33% are Latinx, and 29% are Black. Among the 73 teachers, 30% are male, 39% are Black, 10% are Latinx, averaging 11.3 years of experience.

classroom environment. We then transcribed and coded a random 15-minute interval within each video resulting in 2,820 utterances coded and checked by an expert coder.

Our coding scheme was developed across multiple rounds of coding and discussion and involves the following sets of codes:

a. *Public Praise* (calling out an individual, group of students, or entire class for ideal or desirable behavior) vs. *Public Admonishments* (calling out an individual, group of students, or entire class for disruptive, inappropriate, or undesirable behavior)

b. *Autonomy Supportive Language* (providing students with a choice between activities or strategies) vs. *Controlling Language* (emphasizing the lack of opportunity for autonomy and choice)

c. *Strategy Suggestions* (sharing techniques, tools, or tips for learning and understanding material) vs. *Lack of Strategy Suggestions* (failing to provide a concrete suggestion)

d. *Mindset Supportive Language* (explicitly supporting growth mindset, purpose and relevance, and social belonging) vs. *Mindset Undermining Language* (explicitly undermining growth mindset, purpose and relevance, and social belonging)

Key Findings

We identified a set of prevalent teacher discourse practices likely to affect students' psychological and academic outcomes. These practices varied substantially from lesson to lesson and from teacher to teacher.

There were an average of 20 teacher discourse events in each 15-minute video segment. We found that both *Public Praise* and *Public Admonishment* occurred frequently within mathematics lessons, as did *Strategy Suggestions. Autonomy Supportive Language* (and the opposite, *Controlling Language*) was less frequent but still found at least once in the average 15-minute segment, as were *Mindset Supportive Language. Mindset Undermining Language* occurred infrequently.

Specific findings with respect to the ratio of positive to negative discourse include:

- Teachers used *Public Admonishments* somewhat more often than *Public Praise* (about 44% of evaluative utterances were praise as opposed to admonishment).
- Teachers were substantially more likely to use *Mindset Supportive Language* than *Mindset Undermining Language* (about 82% of mindset-related discourse was mindset supportive).
- Teachers nearly always offered explicit *Strategy Suggestions* (comprising 91% of strategy-related discourse) as opposed to an obvious or complete *Lack of Strategy Suggestions.*
- *Autonomy Supportive Language* (46%) and *Controlling Language* occurred in about equal measure.

We trained observers to achieve high levels of reliability in identifying these discourse dimensions from classroom video and investigated the feasibility of automatic discourse coding to scale up similar analyses.

We found that trained observers could reliably code these focal discourse practices at a fine-grained level: observers coded an average of approximately 20 teacher discourse events (i.e., teacher turns) in each 15-minute video segment. In these events, there is very high agreement on discourse codes, in excess of 95% agreement for the majority of codes.

Preliminary semi-automated (i.e., using transcripts) analyses of teacher discourse indicate moderate performance (mean correlation of 0.55) with human-coded discourse. Correlations were lower (mean of 0.25) but nonzero using nonverbal cues such as paraverbal information (e.g., intonation) and conversational dynamics (e.g., pauses between utterances). These analyses also provided insights into characteristic language patterns accompanying the various discourse categories.

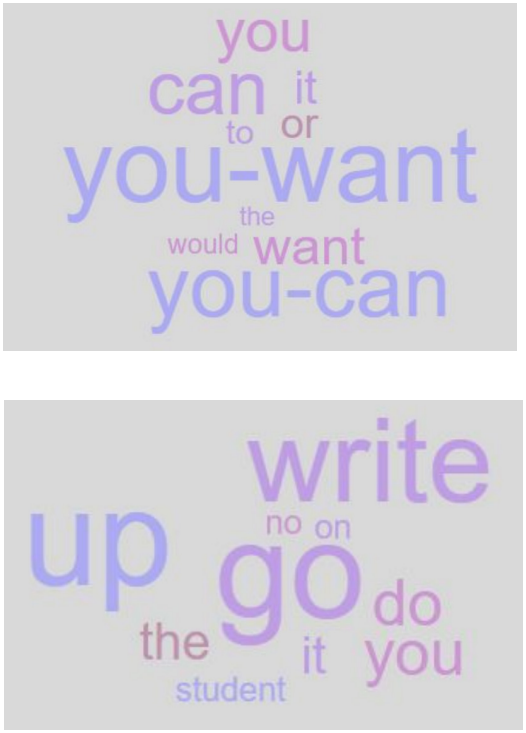

Figure 1. *Word-clouds for* Autonomy Supportive Language *(top) vs.* Controlling Language *(bottom)*

We established the validity of our new fine-grained measures in comparison to existing teacher observation protocols and with student measures of belonging and achievement.

We compared our coding scheme with existing protocols to establish convergent and discriminant validity. Convergent validity can be established by observing a strong relationship between measures that would be expected to overlap theoretically (i.e., they measure related constructs). Discriminant validity can be established by observing little to no relationship between measures that are not expected to overlap theoretically (i.e., they measure distinct constructs).

We found convergence between three of the paired sets of fine-grained discourse measures *(Public Praise/ Public Admonishment, Mindset Supportive Language/ Mindset Undermining Language,* and *Autonomy Supportive Language/Controlling Language*) and observational measures of positive classroom climate and student engagement from the CLASS observation protocol. Demonstrating discriminant validity, we found much lower associations between our codes and the MQI protocol, which focuses more on task complexity and teacher knowledge of mathematical content for teaching, than on the teacher moves that influence motivation and engagement.

We also documented modest but statistically significant relationships between several discourse practices, including *Public Admonishment* and *Autonomy Supportive Language* and student reports (from the TRIPOD measure) of the teacher-student relationship quality as well as happiness in class. *Autonomy Supportive Language* showed the most promising association with mathematics achievement growth. We have yet to document a consistent role of these teacher discourse practices in moderating or mediating gaps in belonging or achievement between student groups.

Insights & Future Directions

There is growing evidence that teachers are eager to engage with feedback on instruction, particularly if this feedback is fine-grained and lesson-specific. Fine-grained measures of relevant discourse can direct attention to teaching practices that facilitate motivation and engagement among all students, and particularly among students who are often marginalized in mathematics classes.

Our work, utilizing a corpus of expert-coded data, helped identify how teachers' discourse practices were linked to both expert observers' and students' perceptions of the classroom environment. Importantly, each of the commonly occurring codes was highly variable from lesson to lesson within the same teacher, suggesting the potential for teacher learning. Similarly, a great deal of the total variance in teacher discourse codes (more than 40% in some cases) was between teachers. Positive practices like *Public Praise, Autonomy Supportive Language, Strategy Suggestions,* and *Mindset Supportive Language* did not necessarily co-occur consistently; a lesson might score high in one domain but average or low in others (and the same is true for discourse occurrences with a negative valence). These findings demonstrate that teachers have considerable potential for growth in how they communicate with their students and create a normative environment of inclusion and learning.

We are exploring how novel technologies can provide teachers with automated feedback on their own discourse to enable a form of data-driven reflective practice and job-embedded professional development. We are optimistic that this work will not only speak to instructional improvement efforts locally, in school and district professional development contexts, but also help direct research efforts towards increasingly fine-grained assessments of instruction.