Statistical Interpretive Discourse Analysis in Educational Research: A Mixed Methods Approach to a Wicked, Complex Problem

Ming Ming CHIU
The Education University of Hong Kong

Mary McVee and Tyler Rinker
University at Buffalo

ABSTRACT
Rather than simply choosing either a qualitative or quantitative approach for analyzing discourse, we propose cycles of interpretive discourse analysis (IDA) and statistical discourse analysis (SDA), namely statistical interpretive discourse analysis (SIDA), a holistic and flexible mixed methods approach. We illustrate SIDA with 3,214 turns of talk by 20 tetrads (80 students) collaborating to solve an algebra problem. Using two SIDA cycles (IDA→SDA→IDA→SDA), we examined how explanatory factors at different levels (groups, time periods, and recent turns) affected the likelihood of a correct, new idea (micro-creativity) in each turn of talk. Specifically, the results showed (a) group differences, (b) summaries were often pivotal moments that ignited a new time period with much more micro-creativity, (c) summary time periods, polite disagreements, and wrong ideas were linked to greater micro-creativity, and (d) rude disagreements were linked to less micro-creativity. SIDA’s complementary cycles of IDA + SDA informed each other’s subsequent analyses and provided mutually supportive evidence. Furthermore, we showed SIDA’s usefulness for analyzing large data sets and how interpretative and statistical analyses complement one another’s strengths and address another’s weaknesses.

Keywords: Discourse analysis, Mixed methods, Education, Conversation, Time-series analysis, Multilevel analysis, Hierarchical linear modeling

Please direct correspondences regarding this monograph to Ming Ming Chiu, Assessment Research Center, Block B1, 2nd floor, Room 15, The Education University of Hong Kong. Email: mingmingchiu@gmail.com. The ideas in this manuscript were inspired by a series of Productive Multivocality workshops organized by Dan Suthers, Kris Lund, Carolyn Rose, Chris Teplovs, and Nancy Law, along with discussions with them, Hajime Shirouzu and Stefan Trausan-Matu. We appreciate the research assistance of Yik Ting Choi.

Please cite as follows:
Statistical Interpretive Discourse Analysis in Educational Research: A Mixed Methods Approach to a Wicked, Complex Problem

The study of how students learn during classroom discourse is both a wicked problem and a complex system. Like other wicked problems, classroom discourse is not easily defined, understood, or accepted (Rittel & Webber, 1973), “involving multiple interacting systems, replete with social and institutional uncertainties, and for which only imperfect knowledge about their nature and solutions exist” (Mertens, 2015, p. 3). Like other complex systems, classroom discourse involves “aspects of the real world for which events and actions have multiple causes and consequences, and where order and structure coexist at many different scales of time, space, and organization” (Jacobson & Wilensky, 2006, p. 12). Classroom discourses occur in complex, dynamic environments influenced by both internal social actors (e.g., teachers and students) and outside social forces and actors (e.g., socio-economic class, stakeholders, policymakers, and historical contexts).

Past studies examined classroom discourse with various methods involving quantitative data (often from close-ended questions or items) or qualitative data (often from open-ended queries). In these studies, researchers sought to understand sequences of social and cognitive processes during face-to-face and online interactions, such as (a) triads of Dutch third-grade students discussing how to write their reports (sequences of student actions linked to high cognition or low cognition, Molenaar & Chiu, 2014), (b) groups of United States teachers analyzing videotaped lessons (sequences linked to new ideas for their teaching; Arya, Christ & Chiu, 2015), and (c) Canadian university student teams designing lessons online (sequences linked to new ideas or theorizing; Chiu & Fujita, 2014).

Hence, a key challenge is how to integrate qualitative and quantitative methods to understand classroom discourse. Past articles discussed how pairs of qualitative and quantitative approaches differ and might complement each other by triangulating their results (e.g., Bangerter & Cornelissen, 2017; De Ruiter & Albert, 2017). In this study, we go further by showcasing how to use cycles of a qualitative analysis (interpretive discourse analysis, IDA) and a quantitative analysis (statistical discourse analysis, SDA), so that the results of one analysis also informs the next analysis. IDA has several core assumptions (e.g., talk helps constitute the social world, is situated within a context, and aims to affect an audience, Woofitt, 2005), and researchers can draw from multiple approaches to analyses (e.g., critical discourse analysis, conversation analysis, ethnography, etc.). Meanwhile, SDA combines multilevel analysis and time-series analysis of turns of talk (Chiu, 2008). One or more cycles of IDA + SDA comprise statistical interpretive discourse analysis (SIDA, see Figure 1). The result of each component analysis (IDA or SDA) not only triangulates each other’s findings but also enhances the next part of the SIDA cycle; the IDA results inform the next SDA model, and the SDA results inform the focus of the next IDA.
A mixed methods approach might help researchers adopt a dialectical view that weaves together qualitative and quantitative approaches. We consider examples of qualitative and quantitative discourse analyses that highlight the challenges and benefits of mixed methods research (MMR) for studying discourse analysis. Then, we then present SIDA in detail via a collaborative problem solving study of 80 students in 20 tetrads working on an algebra problem during one classroom lesson (Chiu, 2008). We conclude by discussing SIDA’s affordances and limitations.

**Combining Qualitative and Quantitative Approaches in Discourse Analysis**

Integrating MMR in discourse analysis is rare (Bryman 2006; of all *Journal of Mixed Methods Research* articles since 2007, only two mentioned both qualitative and quantitative methods for discourse analysis, but neither addressed MMR methodological issues). Discourse studies predominantly use either (a) quantitative methods in a positivist or postpositivist paradigm of testing hypotheses regarding outcomes, possibly with interpretive approaches to identify categories (Graesser et al., 1997; Diehl & McFarland, 2012; Teasley, 1995; Wolfe et al., 1998) or (b) qualitative methods in an ethnographic or interpretive paradigm to analyze processes, possibly quantifying data with descriptive statistics (e.g., frequencies or percentages; Green & Dixon 2002). Researchers rarely combine qualitative and quantitative methods on equal footing in a study.
A mixed methods approach might help researchers adopt a dialectical view that weaves together qualitative and quantitative approaches. Viewing discourse as objective, reductive and discrete, past quantitative studies of educational discourse have focused on performance capacity to predict or show how well people perform on particular tasks under specific conditions (Harré & Moghaddam 2003). For example, Chiu (2018) showed that during collaborative problem solving of a mathematics problem, students who created more new ideas that were correct (micro-creativity), evaluated ideas correctly, or had more sequences of a correct evaluation followed by micro-creativity (correct evaluation \([1 \rightarrow \text{micro-creativity}]\) were more likely to correctly solve an algebra problem (via multilevel analysis and time-series analysis). Such studies assume that attribute categories for turns of talk (or other units of time), such as micro-creativity, capture sufficient information from the underlying phenomena (at multiple levels of turn of talk, time period, individual, group, school, etc.) to make inferences (Chiu & Lehmann-Willenbrock, 2016). These studies also assume that instances of categories are sufficiently similar to be treated as equivalent (reliability) for (a) the quantitative analysis and (b) related inferences to be generalizable to other such instances in other contexts (validity, Chiu, 2013).

In contrast, qualitative educational researchers often examine performance style (Harré & Moghaddam 2003) in discourse with constructivist or transformative worldviews (including critical discourse analysis, Fairclough 2010, see Figure 2). Qualitative studies of discourse can examine socially-situated, meaning-making, language-in-use (little ‘d’ discourse) and/or shaping of learner identities in social and historical contexts (big ‘D’ Discourse, Gee, 2014). Consider Cazden’s (2001) classic example of how elementary school students’ stories during sharing time reflected their home discourse. Whereas a teacher appreciated White students’ topical narratives (tightly organized in a linear arc to support and develop a story topic), she stopped Black children from sharing episodic narratives (with its shifting scenes, characters, and potentially non-linear storylines), which to her seemed incoherent with unrelated and unimportant details. Believing that these Black students could not form coherent stories, a teacher can place them in lower reading/writing groups for lacking performance style rather than performance capacity.

As qualitative discourse analysts tend to “view discourse as both constitutive of social worlds and as shaped by those social worlds” (p. 18) rather than reductionist and discrete, Bryman (2007) argued that those using discursive approaches might face ontological challenges in mixed methods research. Viewing qualitative and quantitative paradigms as mutually exclusive, Glesne (2011) claimed: “If you were to attempt to combine positivist methodology that relies heavily on quantitative methods such as experimental design with an interpretivist methodology that relies on qualitative methods such as ethnography, you would end up doing two studies” (p. 14).

In contrast, Mertens (2012) proposes a dialectic paradigm:

This stance allows the researcher to adhere to the beliefs of the postpositivist paradigm in conducting quantitative-oriented data collection and of the constructivist paradigm in qualitative-oriented data collection and then to put the two in conversation with each other throughout the study to allow for deeper
understandings based on the convergence and dissonance found in the approaches (p. 256).

Embracing this dialectic stance (see Figure 2), SIDA “actively welcomes more than one paradigmatic tradition and mental model, along with more than one methodology and type of method, into the same inquiry space and engages them in respectful dialogue” (Greene & Hall 2010, p. 124).

![Diagram showing four worldviews that can inform SIDA](image)

Figure 2. Four worldviews that can inform SIDA (adapted from Creswell, 2014, p. 6).

SIDA interweaves performance capacity and performance style strands to mutually support and inform one another (see Figure 1), thereby addressing “a broader and more complete range of research questions” than any single method can (Johnson & Onwuegbuzie, 2004, p. 21). Quantitative analysis of discourse allows the analyst to examine large data sets, control for the simultaneous effects of multiple variables, and establish reliability, validity, and generalizability (Creswell, 2015). Specific quantitative techniques such as regression and SDA have further benefits. The unexplained components (residuals) of regressions identify specific data points that fit the theoretical model well (prototypical instances with small residuals) versus poorly (exceptions with large residuals, Kennedy, 2008). Furthermore, SDA statistically identifies the pivotal moments that radically change an interaction (Chiu & Lehmann-Willenbrock, 2016). Complementing these statistical analyses, qualitative analyses can detect key differences among instances within a category (e.g., disagreements: polite vs. rude), illuminate the contextual meaning of quantitative relationships (why do disagreements often precede new ideas?), examine instances in detail (e.g., prototypical instances, exceptions,
watersheds), and identify other crucial aspects outside quantitative categories (e.g., activity frames, Mercer et al., 2004).

In the next section, we turn our attention to the SIDA model. We describe and illustrate IDA, SDA, and their cycles of analyses.

**An Illustration of SIDA via the IDA <= SDA Cycle**

In this section, we illustrate how SIDA cycles of IDA and SDA (see Figure 1) can inform each subsequent analysis, generate and test hypotheses more precisely, and yield additional insights. We introduce the research question and data, before applying SIDA.

**Changing the Research Question**

Initially, the lead author Lane (a pseudonym for blind review) planned to do a dissertation using qualitative methods to examine metaphorical understandings of mathematics (e.g., variables are containers, Lakoff & Nunez, 2013) by groups of students working on an algebra problem. He chose to examine groups solving algebra word problems for three reasons. First, as its solution required multiple ideas and computational steps, it was sufficiently difficult and complex to elicit metaphorical thinking (Chiu, 2001). Second, he could easily evaluate the validity of each mathematics idea from each student. Third, the actions and talk by groups of students while solving this problem could show evidence of their thinking. Although he found too few metaphor examples to analyze, he did not want to abandon his dissertation data, so he changed his research focus to a different question: what sequences of group processes during collaborative problem solving increase (or decrease) micro-creativity?

**Data**

Eighty students attended four ninth-grade algebra classes in an urban US high school for 7 months. They comprised 40 girls and 40 boys; there were 12 Asians, 27 Blacks, 28 Hispanics, and 13 Whites. These students worked in groups of four (all groups were heterogeneous with respect to gender and race). This class lesson was the first day of both group work and a new unit on algebraic equations with multiple variables. Their teacher asked them to work on the following problem for 30 minutes:

You won a cruise from New York to London, but you arrive 5 hours late. So, the ship left without you. To catch the ship, you rent a helicopter. The ship travels at 22 miles an hour. The helicopter moves at 90 miles an hour. How long will it take you to catch the ship?

One solution equates the distance computations for the cruise and for the helicopter.

\[ 22 \text{ mph} \times [\text{Time} + 5 \text{ hours}] = 90 \text{ mph} \times \text{Time} \]

Isolating \( \text{Time} \) via algebraic operations yields 1.618 hours or 1 hour 37 minutes. Each group’s problem solving was videotaped and transcribed.

**First IDA (Qualitative)**

In the first IDA, Lane sought to understand which sequences of student talk raised micro-creativity via **contrasting cases** (Roelle & Berthold, 2015); he selected and examined two conversations, one group correctly solved the problem and the other did not. By analyzing each group’s talk, he clarified the categories of: correct idea, wrong idea, new idea, old idea, and disagreement.

**Correct ideas and wrong ideas.** Initially, Lane distinguished between mathematically correct and wrong ideas. However, after a student Ana said, “ninety
times fifty is four hundred and fifty,” he realized that some ideas are mathematically correct but do not help solve the problem. Thus, he redefined correct to be consistent with both the problem situation and mathematics. Unlike open problems like writing essays, a simple algebra word problem has a small, finite set of correct ideas that can contribute to a correct solution. Although a task analysis and interviews with students can identify major ideas that contribute to a solution, identifying all possible correct ideas that contribute to a correct solution is not easy or even possible in some cases. Hence, analysis of discourse might proceed with an incomplete map of possible correct ideas.

New ideas and old ideas. Moreover, a correct idea can be new or old. Correct, new ideas (micro-creativity) are often more valuable than correct, old ones. However, distinguishing between new and old is not trivial. For example, a student, Juan said, “five times twenty-two is one ten [5 x 22 = 110].” As no one in this group problem solving session had done this computation, it could be a new idea. However, if Jan computed it months ago for a different problem, it is not new for Jan. As knowing a person’s entire history is practically impossible, Lane compromised by defining a new idea as an expressed idea beyond both the problem statement and past discussion in this activity. While this definition is imperfect, it might be sufficient for this analysis.

Evaluations. Lane also noticed that group members might evaluate ideas correctly or incorrectly. In addition to agreeing with correct ideas, students can also disagree with wrong ideas, which can foster micro-creativity. Consider the following excerpt.

Bob: [hits calculator keys 15 x 50 = 750]
Ben: What is—what is 70,000, 750 mean?
Lex: I think the wrong things got multiplied; try 90 and 1.5.
Bob: [laughs, hits calculator keys 1.5 x 90 = 135] It's 135.

After Bob’s incorrect computation, Ben questions it (“what is 70,000, 750 mean?”). Then, Lex diagnoses the problem (“wrong things got multiplied”) and suggests the correct computation (“try 90 and 1.5”), which Bob calculates correctly (“135”). Hence, disagreeing with wrong ideas can yield micro-creativity. On the other hand, groupmates can also evaluate incorrectly, agreeing with wrong ideas or disagreeing with correct ideas.

In short, the first IDA identified coherent constructs and preliminary relationships among them through recursive analysis of the categories and the transcripts. Ultimately, six constructs were identified through initial qualitative analysis: new ideas, old ideas, correct ideas, wrong ideas, agree, and disagree. Combinations of these categories yielded additional constructs, such as micro-creativity (= new + correct) and correct evaluation (= agree with correct idea OR disagree with wrong idea). The first IDA identified many patterns for the first SDA to test. Due to space considerations however, we focus on whether one sequence wrong idea → disagreements → micro-creativity occurs more often than other sequences.

From Qualitative IDA Ideas to a Quantitative SDA Database
To create a database suitable for statistical analysis, Lane had to identify suitable units of analysis (e.g., utterance, turn, episode, etc.) and categorize his qualitative, video transcript data (Boyatzis 1998). As Lane was interested in how group members affect one
another’s micro-creativity, he chose a turn of talk as the unit of analysis; a turn of talk is a sequence of one person’s words or actions that are bracketed by those of others.

**Categories.** Choosing suitable numbers and types of categories is challenging. While using too few categories can miss important attributes, using too many categories can unnecessarily increase complexity and reduce reliability. Specifically, greater numbers of categories or greater complexity of categories tend to: (a) increase training time for coders, (b) increase overall coding time, (c) reduce internal consistency, (d) increase coding conflicts, (e) reduce inter-coder reliability, (f) reduce degrees of freedom in the explanatory model, and (g) reduce the accuracy of the analyses’ results (Chiu & Khoo, 2005).

As Lane pondered how to categorize such complex behaviors, he chanced upon a three-dimensional category scheme for counseling in an open article on a library table (Stiles 1979); he saw the value of multiple dimensions for reducing coding complexity (and potential for statistical analysis, which led to inventing SDA). By coding one dimension at a time with a decision tree, a coder uses clear criteria to choose among only the possible codes within that dimension rather than all possible codes (Chiu & Khoo, 2005).

Building on IDA’s identification of categories and their complex patterns, Lane created a three-dimensional category scheme (evaluation [agree, disagree, ignore], knowledge [new, old, no content], validity [correct, wrong, no content]), which captures 15 different types of actions (15 = 3 x [2 x 2 + 1] = # of evaluations x [# knowledge types x # invitation types + 1 no content]; Chiu, 2000). Rather than considering 15 possible categories, a coder only considers the 3 possibilities along each dimension. Thus, multi-dimensional coding can (a) allow combinations of simpler explanatory and outcome variables to capture complex phenomena, (b) reduce the number of variables, (c) reduce training time, (d) allow many coders to work simultaneously, (e) reduce coding time, (f) increase degrees of freedom in the explanatory model, and (g) increase the accuracy of the analyses’ results (Chiu & Khoo, 2005).

**Decision trees.** To help coders place each datum in the appropriate category, Lane created decision trees based on the first IDA. Good decision trees delineate boundaries between categories, reduce the number of decisions, and reduce training time. To achieve these aims, a decision tree uses yes/no questions to identify easier, distinct categories before using finer, complex criteria for difficult categorizations. For example, when distinguishing among a new idea, an old idea, and no content (e.g., off-task), he first identified no content turns by asking “does the speaker express any mathematics or problem-related information?” To distinguish between new versus old ideas in the remaining turns, he created a log/trace of new ideas expressed by a group of students and checked each turn for unlisted ideas to add to the log (see Figure 3). As identifying an old idea does not require reading the entire log, it takes less time than identifying a new idea. Hence, the decision tree identifies an old idea before a new idea.
Information gathered from coding with earlier, skeletal decision trees can inform later codes, thereby enhancing inter-rater reliability, reducing effort, and saving time. For example, turns with no content cannot be categorized as correct or wrong, so they were all coded as “no content,” which saved Lane a lot of time. To facilitate coding correctness, Lane made the easier decision about mathematical correctness first and then checked for consistency with the problem situation later (see Figure 4). To identify micro-creativity, he wrote a computer program/macro (if [new and correct] then micro-creativity = true; else micro-creativity = false).

![Figure 3. Decision tree for coding new idea, old idea and null content](image)

![Figure 4. Decision tree for coding correct, wrong and null content](image)
Next, Lane created a decision tree for agreeing, disagreeing, and unresponsive/ignore (see Figure 5). To create a sharp boundary between agreeing and disagreeing, he only coded turns that fully agreed with the previous speaker as agree; even the smallest disagreement was coded as disagree.

Does the current speaker respond to the previous speaker’s idea?

Yes

No

Does he speaker fully agree with the previous speaker?

Yes

Unresponsive / Ignore

No

Agree

Disagree

Figure 5. Decision tree for coding agree, disagree and unresponsive/ignore

**Inter-rater reliability.** Good decision trees based on the first IDA facilitate accurate coding and high inter-rater reliability. To measure inter-rater reliability, Lane used Krippendorff’s (2012) $\alpha$, which is applicable to incomplete data, any sample size, any measurement level, any number of coders or categories, and scale values. Ranging from -1 to 1, an $\alpha$ exceeding .80 shows satisfactory agreement (Krippendorff, 2012). The high inter-rater reliabilities of these data (.95—.99; see Table 1) yield smaller measurement errors and greater accuracy in the analyses’ results (Chiu & Khoo, 2005).

<table>
<thead>
<tr>
<th>Coding dimension</th>
<th>Agreement %</th>
<th>Krippendorf's $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of previous action</td>
<td>97</td>
<td>.95</td>
</tr>
<tr>
<td>Knowledge content</td>
<td>98</td>
<td>.98</td>
</tr>
<tr>
<td>Correct idea</td>
<td>99</td>
<td>.99</td>
</tr>
</tbody>
</table>

**Using SDA to Address Major Analytic Challenges**

Statistically analyzing antecedents of micro-creativity across time (e.g., sequences of individual actions and social interactions) requires overcoming several difficulties regarding the data set, target processes, and explanatory model. As simpler analyses (e.g., ordinary least squares regressions) do not address these issues, they often yield biased results (Kennedy, 2008). SDA addresses these difficulties with several analytic strategies, which can be integrated or used separately (see Table 2).
Table 2
*Addressing a Subset of Analytic Difficulties with Some Statistical Discourse Analysis Strategies*

<table>
<thead>
<tr>
<th>Analytic difficulty</th>
<th>Statistical Discourse Analysis Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole data set</td>
<td>• Use small unit of analysis to raise sample size</td>
</tr>
<tr>
<td>• Minimum sample size</td>
<td></td>
</tr>
<tr>
<td>• Missing data (0110??10)</td>
<td>• Markov Chain Monte Carlo multiple imputation</td>
</tr>
<tr>
<td>• Parallel conversations</td>
<td>• Store turn of talk to which a turn of talk responds</td>
</tr>
<tr>
<td>• Complex categories/codes (A, B…Q)</td>
<td>• Multi-dimensional decision trees</td>
</tr>
<tr>
<td>Dependent variables</td>
<td></td>
</tr>
<tr>
<td>• Differences across time periods (T1≠T2)</td>
<td>• Breakpoint analysis &amp; Multilevel analysis</td>
</tr>
<tr>
<td>• Nested data (turns of talk within time</td>
<td>• Multilevel analysis</td>
</tr>
<tr>
<td>periods within groups)</td>
<td></td>
</tr>
<tr>
<td>• Serial correlation (t3 similar to t4)</td>
<td>• Q-statistics</td>
</tr>
<tr>
<td>• Discrete (yes/no)</td>
<td>• Logit / Probit</td>
</tr>
<tr>
<td>• Infrequent (000010)</td>
<td>• Logit bias estimator</td>
</tr>
<tr>
<td>Explanatory variables</td>
<td></td>
</tr>
<tr>
<td>• Sequences (X_{t-2} or X_{t-1} → Y_0)</td>
<td>• Vector Auto-Regression</td>
</tr>
<tr>
<td>• Indirect, mediation effects (X → M → Y)</td>
<td>• Random effects model</td>
</tr>
<tr>
<td>• False positives (Type I errors)</td>
<td>• Multi-level M-tests</td>
</tr>
<tr>
<td>• Robustness of results</td>
<td>• Two-stage linear step-up procedure</td>
</tr>
<tr>
<td></td>
<td>• Analyses of subsets of data</td>
</tr>
<tr>
<td></td>
<td>• Analyses of original, unestimated data</td>
</tr>
</tbody>
</table>

**Data issues.** The utility of a statistical analysis depends on the quality of the data; major data issues include sufficient sample size, missing data, and parallel conversations. As sufficient *statistical power* requires a large sample size, using a smaller unit of analysis (turns rather than groups) increases the sample size (from 20 groups to 3,214 turns) and decreases the required data collection, time, and costs (e.g., 100 turns from a few groups rather than many more turns from 100 groups; for detailed power analyses, see Konstantopoulos, 2008).

Missing data can (a) bias the results, (b) increase the complexity of the analyses, and (c) reduce estimation efficiency (Peugh & Enders 2004). *Markov Chain Monte Carlo multiple imputation* estimates the values of the missing data, which outperforms *list-wise deletion, pair-wise deletion, mean substitution*, and *simple imputation* according to computer simulations (Peugh & Enders 2004).
Also, conversations do not always proceed neatly with one turn following another (e.g., group discussions can split into parallel conversations in which turn 4 responds to turn 2 [not turn 3]). To address this issue, we identify and store the prior turn to which the current turn responds. Then, we create variables indicating values of previous turns (i.e., lag variables, e.g., wrong \([t-1]\) indicates whether the previous turn had a wrong idea). Hence, these variables can capture parallel conversations.

**Target process issues.** Issues regarding target processes (e.g., micro-creativity) include differences across groups, differences across time, dichotomous outcomes, and infrequent outcomes. As students taught by the same teacher likely resemble one another more than those taught by different teachers (nested data), an ordinary least squares regression biases the standard errors, so we use a multilevel analysis (Goldstein, 2011; also known as hierarchical linear modeling, Bryk & Raudenbush, 1992).

Processes can also differ across time. Failure to account for similarities in turns of talk within the same time period or in adjacent turns (serial correlation of residuals) can bias the standard errors (Kennedy, 2008). Breakpoint analysis statistically identifies pivotal moments that distinguish time periods of high versus low frequencies of target processes (micro-creativity) and models them with a time-period level in a multilevel analysis (Wise & Chiu, 2011). \(Q\)-statistics test each group for serial correlation in adjacent turns (Ljung & Box, 1979). If serial correlation of the target process (micro-creativity) is significant, adding the lagged outcome variable micro-creativity \([t-1]\) as an explanatory variable often removes the serial correlation (Chiu & Khoo, 2005). These time periods are another level of analysis (in addition to class, group, and turn of talk).

For dichotomous outcomes (e.g., micro-creativity vs. no micro-creativity in each turn), ordinary least squares regressions can bias the standard errors. Thus, we use a Logit regression to model dichotomous outcomes correctly (Kennedy, 2008). To aid understanding of these logit results, we compute the odds ratio of the regression coefficient and report it as the percentage increase or decrease in the likelihood of the outcome (Kennedy, 2008). Infrequent events (occurring in less than 25% of the data) can bias logit regression results, so we estimate the bias and remove it (King & Zeng, 2001).

**Explanatory model issues.** Explanatory variable issues include sequences, indirect effects, interactions across levels, many hypotheses’ false positives, comparison of effect sizes, and robustness. As preceding turns might influence the current turn, the analysis must model previous sequences of turns (Kennedy, 2008). A vector auto-regression (VAR, Kennedy, 2008) tests whether attributes of sequences of recent turns (micro-time context) influence the current turn (e.g., the likelihood of micro-creativity). Also, separate, single-level tests of indirect mediation effects on nested data can bias results, so we test for indirect multi-level effects with a multilevel M-test (MacKinnon et al., 2004).

For these nested data, modeling interaction effects across levels (e.g., student x turn of talk) with a fixed effects model can bias the results (Goldstein, 2011), so we use a random effects model (Goldstein, 2011). If the regression coefficient of an explanatory variable (e.g., \(\beta_{yv} = \beta_{yv0} + f_{yv}\)) differs significantly across levels (\(f_{yv} \neq 0\)), then
Cross-level moderation might exist, and we model the regression coefficient with structural variables (e.g., student gender).

As testing many hypotheses increases the possibility of a false positive, we reduce their likelihood via the two-stage linear step-up procedure, which outperformed 13 other methods in computer simulations (Benjamini et al., 2006).

When testing whether explanatory variables have different effect sizes, Wald and likelihood ratio tests do not apply at boundary points. Hence, we use Lagrange multiplier tests which apply to all parts of a data set and show greater statistical power than Wald or likelihood ratio tests for small deviations from the null hypothesis (Bertsekas, 2014).

We use two variations of the core explanatory model to test whether the results remain stable despite minor changes in the data or analysis (robustness, Kennedy, 2008). We run subsets of the data separately to test the consistency of the results across subsets. Then, we repeat the analyses for the original, un-estimated data.

Like traditional regressions, SDA assumes a linear combination of normally distributed explanatory variables. (Nonlinear aspects can be modeled as nonlinear functions of variables [e.g., \((\text{mathematics grade})^2\)] or interactions [e.g., disagree x wrong -1].) SDA also requires independent and identically distributed residuals, and a modest minimum sample size. (As SDA uses a multilevel analysis, it does not require homoscedasticity [Goldstein, 2011].)

**First SDA (Quantitative)**

**Analysis specification.** After the first IDA identified coherent constructs and possible relations, the first SDA tests these relations, notably whether wrong ideas or disagreements increase micro-creativity. Lane estimates the missing data, statistically identifies breakpoints and time periods, and then tests an explanatory model. Estimation of the missing data (2%) yielded complete data for all groups.

*Figure 6.* Statistically-identified breakpoints divided the data into 2 time periods in the unsuccessful group and 5 time periods in the successful group.

Next, Lane statistically identified the breakpoints that divide each group’s problem solving session into time periods with significantly larger versus smaller likelihoods of micro-creativity. For example, the unsuccessful group in Figure 6 has one breakpoint (two time periods), whereas the successful group has four breakpoints (five
time periods). These statistically identified breakpoints likely affected micro-creativity for substantial portions of time and pinpoint pivotal moments for further IDA.

However, SDA’s multilevel analysis heuristically requires 20 units at the highest level to converge to a solution (Chiu & Khoo, 2005), far more than the current two groups and seven time periods. Hence, a simpler version of SDA—a single-level logit analysis was used.

\[
\pi = p(\text{micro-creativity}_i = 1 \mid \beta_0) = F(\beta_0) \tag{1}
\]

In the basic model, \( \pi \) is the probability that each turn \( i \) has \textit{micro-creativity}, with an intercept \( \beta_0 \). Next, explanatory variables are added. \textit{Group}_2 tests whether it shows more (or less) \textit{micro-creativity} than \textit{Group} 1. Then, student demographics (\textit{girl}, \textit{Black}), past achievement (\textit{mid-year algebra grade}) and action (\textit{disagree}) (\textit{Current_Speaker}) are tested for their relation to \textit{micro-creativity}. Next, the previous speaker’s demographics, past achievement, and actions (\textit{wrong idea}-1) are added (\textit{Previous_Speaker}).

\[
\pi = F(\beta_0 + \beta_1 \text{Group}_2 + \beta_c \text{Current_Speaker}_i + \beta_p \text{Previous_Speaker}_{i-1}) \tag{2}
\]

**Results.** The explanatory model results showed differences across groups and links between attributes of recent turns to \textit{micro-creativity} in the current turn of talk (see Table 3). \textit{Group} 2 students were 35% less likely than \textit{Group} 1 students to show \textit{micro-creativity} (35% = odds ratio of -1.748, see Table 3, model 1, top, Kennedy 2008). After a wrong idea (-1) by the previous speaker, \textit{micro-creativity} was 42% more likely (42% = odds ratio of 2.459, see Table 3, model 3, bottom). All other explanatory variables were not significant.

Notably in model 2, \textit{disagree} had a large positive regression coefficient (0.91), a large standard error (0.54), and large differences in its residuals—many large (poor fit) and many small (good fit). These results suggest that examining various disagreements closely might distinguish one type that increases \textit{micro-creativity} from another type that reduces it.

Overall, the first SDA’s results pinpoint three subsets of data for the second IDA to examine in greater detail: breakpoints, best-fitting turns, and worst-fitting turns.

**Second IDA of the first SDA’s identified key moments**

Our second IDA examines specific moments identified by the first SDA: (a) breakpoints to reveal how they radically change each group’s interactions and \textit{micro-creativity}, (b) best-fitting turns involving wrong idea (-1), \textit{disagree}, and \textit{micro-creativity} to detail the theoretical model’s mechanisms, and (c) worst-fitting turns with \textit{disagree} to explore different mechanisms.

**Breakpoints.** IDA of the breakpoints yields three major categories: \textit{transitions between on-task and off-task behaviors}, \textit{insights}, and \textit{error spawners}. The transitions consisted of individuals initiating \textit{off-task topics} (“We went to Maryland last summer”) or \textit{on-task topics} (“What’s the distance between New York and London?”) and groupmates responding to them.
Table 3  
Summary of Standardized Regression Coefficients (and Their Standard Errors) in the First SDA Results Modeling Micro-Creativity in Two Groups (n = 108 Turns of Talk)  

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Model 1 Group</th>
<th>Model 2 + Current speaker</th>
<th>Model 3 + Previous speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.748 ***</td>
<td>-1.696</td>
<td>-7.044 *</td>
</tr>
<tr>
<td></td>
<td>(0.428)</td>
<td>(1.202)</td>
<td>(3.094)</td>
</tr>
<tr>
<td>Black a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.508</td>
<td>-2.618</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.848)</td>
<td>(1.512)</td>
<td></td>
</tr>
<tr>
<td>Girl</td>
<td>-0.692</td>
<td>-0.971</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.858)</td>
<td>(1.406)</td>
<td></td>
</tr>
<tr>
<td>Math grade</td>
<td>0.038</td>
<td>-0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.079)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>0.912</td>
<td>0.300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.538)</td>
<td>(0.617)</td>
<td></td>
</tr>
<tr>
<td>Black (-1)</td>
<td></td>
<td></td>
<td>-0.933</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.399)</td>
</tr>
<tr>
<td>Girl (-1)</td>
<td></td>
<td></td>
<td>-3.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.753)</td>
</tr>
<tr>
<td>Math grade (-1)</td>
<td></td>
<td></td>
<td>-0.156</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.087)</td>
</tr>
<tr>
<td>Wrong idea (-1)</td>
<td></td>
<td></td>
<td>2.459 **</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.816)</td>
</tr>
<tr>
<td>Explained variance</td>
<td>0.124</td>
<td>0.181</td>
<td>0.298</td>
</tr>
</tbody>
</table>

Note. A constant term was included. * p < .05; ** p < .01; *** p < .001  
a Both groups only had Black and Latino students.

Many insight breakpoints were summaries of the group’s previous discussion (cf. Wise & Chiu, 2011). In the following example, a student drew a diagram that summarized their progress [breakpoint in italics].

Liz It's 5 hours, so [hits calculator keys, 90 x 5] 450?
Amy [draws] Ok, this is like. Okay, so like, ok, this [points to drawing] is New York, right? And that's London [points to drawing].
Liz Right.
Amy [pointing to drawing] Okay, okay um, that’s the cruise ship. Ok. And like the cruise ship is ahead of the helicopter, right?
Liz Yeah. At a 110.
Amy Okay, [writes 110 near the cruise ship symbol], the helicopter's moving up.
Max That's a helicopter?
Amy Well, [raises open hands] I can't draw [laughs]
Max [laughs] Alright.
Amy Okay? You've got to think about the time. We have 90—
Liz — We have this um to deal with this [points to cruise ship on drawing] too because it's not gonna stop.

Amy Oh, the cruise ship's not gonna stop.

Students (e.g., Liz) added and multiplied numbers from the problem, making little progress. Then, at the breakpoint, Amy correctly diagramed three old ideas about the problem situation: the ship’s location, the helicopter’s location, and the helicopter’s movement. Liz elaborated this summary with the ship’s distance from shore ("110"). After some friendly teasing about the poor drawing, Amy highlighted the helicopter’s travel time, and Liz interrupted to point out the ship’s continuing motion ("it's not gonna stop"). Amy validated Liz's idea, and the group then computed each vehicle’s movements, marked them on the diagram, and solved the problem. In short, the diagram breakpoint elevated the group’s micro-creativity as they solved the problem.

Other breakpoints spawned errors. For example, after several correct computations, a student mistakenly divided the helicopter speed by the ship distance. His groupmates did not notice this error and used the arithmetic result in further calculations, spawning more wrong ideas. As a result, this group did not solve the problem.

In short, breakpoints included transitions between on-task versus off-task, insights, and error spawners. Notably, summaries were readily identifiable as containing three or more old ideas, and can enhance the explanatory model in the second SDA.


Max: 5 plus 22.
Liz: We need to do time times rate, so 5 times 22.
Max: [presses calculator keys] 110.

When Liz politely disagrees with Max’s wrong idea, she highlights their common cause through shared positioning ("we") and justifies ("time times rate") the change in computation ("5 times 22") to help him preserve his public self-image (save face). Then, Max agrees and responds with micro-creativity ("110").

In contrast, the worst-fitting turns differ substantially.

Jim: Six times ninety.
Lex: No, that doesn’t make sense.
Jim: Yes, it does.

Unlike Liz, Lex rudely disagrees; he rejects Jim’s suggested computation ("no") and denigrates it ("that doesn’t make sense"), both of which threaten Jim’s face. In response, Jim rejects Lex’s evaluation and re-asserts his proposal ("Yes, it does"). The prevalence of polite disagreements in well-fitting turns and of rude disagreements in poorly-fitting turns suggest that distinguishing between polite and rude disagreements might improve (a) our explanatory model of micro-creativity and (b) the model’s fit with the data.

Compared to the first IDA+SDA cycle, the second IDA suggests a superior explanatory model of four additional hypotheses. First, summaries are more likely than other turns to yield breakpoints. Second, polite disagreements are more likely than other turns to be followed by micro-creativity. Third, wrong ideas show an indirect effect on micro-creativity via polite disagreement (wrong idea → politely disagree → micro-creativity). Lastly, rude disagreements are less likely to precede micro-creativity.
Second SDA Tests Second IDA’s Hypotheses

Unlike the first SDA on two groups, the second SDA tests the second IDA’s new hypotheses on all groups. All data were coded for summaries, polite disagreements, and rude disagreements. Then, SDA identified breakpoints for each group.

Next, Lane created decision trees for summary, politely disagree, and rudely disagree. Defining a summary as three or more old ideas seemed to cover the turns that Lane marked as summaries in the second IDA. Also, as group members sometimes collaborated on a joint summary across several turns, the decision tree defines a summary as occurring within five turns (allowing for backchannel acknowledgements like “uh-huh”). See decision tree in Figure 7.

Then, disagree turns were divided into rude disagreements and polite disagreements (see Figure 8). If a disagree turn included any of the politeness strategies in Table 4 (Chiu, 2008), it was coded as politely disagree; otherwise, it was coded rudely disagree (when disagreeing, using a politeness strategy is the norm; thus, its omission is considered rude, Holtgraves 1997).

Table 4
Politeness Markers

<table>
<thead>
<tr>
<th>Politeness markers</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive evaluation</td>
<td>Excellent!</td>
</tr>
<tr>
<td>Question</td>
<td>Should we do this?</td>
</tr>
<tr>
<td>Shared references to audience and self</td>
<td>Ana and I, we, our</td>
</tr>
<tr>
<td>Passive voice</td>
<td>Is multiplied</td>
</tr>
<tr>
<td>Passive verbs</td>
<td>Get, have</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Perhaps, maybe</td>
</tr>
<tr>
<td>Generalized subject</td>
<td>People often …</td>
</tr>
<tr>
<td>Double negatives</td>
<td>not wrong</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>If … ; let’s say …</td>
</tr>
<tr>
<td>Offers (self-reference with dynamic modal)</td>
<td>I can multiply them</td>
</tr>
<tr>
<td>Apology</td>
<td>I’m sorry</td>
</tr>
</tbody>
</table>

Note. Politeness markers are sorted from most common (top) to least common (bottom) to reduce coding time.

Number of Breakpoints. The number of breakpoints per group ranged from zero to five, yielding one to six time periods, respectively, for a total of 52 breakpoints and 72 time periods. Twenty-six breakpoints were transitions between on-task versus off-task (8 in successful groups vs. 18 in unsuccessful groups). 14 breakpoints (8 vs. 6) showed summaries or insights, yielding more micro-creativity afterwards. The remaining 12 breakpoints (7 vs. 5) started cascades of wrong ideas.
Figure 7. Decision tree for coding summary
Two separate SDA analyses were run. The first analysis tested whether summaries were more likely than other turns to be breakpoints.  

$$\pi_{ijk} = p(Breakpoint_{ijk} = 1 \mid \beta_0) = F(\beta_0)$$  
(3)

The outcome $\pi$ is the probability that a breakpoint occurs at turn $i$ in time period $j$ in group $k$. Then, the explanatory variable $\text{Summary}$ is added.  

$$\pi_{ijk} = F(\beta_0 + \beta_1 \text{Summary}_{ijk})$$  
(4)

With 20 groups, 72 time periods, and 3,214 turns of talk, SDA’s multilevel analyses was used (turns within time periods within groups).

**Breakpoint model results.** The results showed that the differences in breakpoints differed significantly at all three levels: group (30%), time period (28%) and turn (42%). Summaries were 47% more likely than other turns of talk to be breakpoints (47% = odds ratio of 3.399, SE = 0.393) and accounted for 8% of the variance in micro-creativity across the data. These results show that we can explain which actions are more likely than others to be pivotal moments (in this case, summaries) that radically change the interaction.

**Explanatory model of micro-creativity.** The second analysis tested whether wrong ideas, disagreeing politely, or disagreeing rudely were related to micro-creativity.

$$\pi_{ijk} = p(\text{micro-creativity}_{ijk} = 1 \mid \beta_0) = F(\beta_0)$$  
(5)

The outcome, $\pi$, was the probability that micro-creativity occurs at turn $i$ in time period $j$ in group $k$.

$$\pi_{ijk} = F(\beta_0 + \beta_g \text{Group}_k + \beta_1 \text{Summary_time_period}_{jk} + \beta_c \text{Current_Speaker}_{ijk} + \beta_p \text{Previous_Speaker}_{(i-1)jk})$$  
(6)

With many groups, several group variables were tested (group percentages, means or standard deviations of individual variables): girl%, Asian%, Black%, Latino%, mid-year algebra grade_mean, mid-year algebra grade SD (Group). In addition, the time period variable $\text{Summary_time_period}_{jk}$ was tested. $\text{Current_Speaker}$ consisted of girl, Asian, Black, Latino, mid-year algebra grade, politely disagree, and rudely disagree.  

$\text{Previous_Speaker}$ consisted of girl (-1), Asian (-1), Black (-1), Latino (-1), mid-year
algebra grade (-1), and wrong idea (-1).

**Micro-creativity results.** Unlike the first SDA results (see Table 3), summaries, disagreements, and wrong ideas were all related to micro-creativity in this second SDA (see Table 5). In time periods immediately after summary breakpoints, micro-creativity was 31% more likely to occur, compared to other time periods (odds ratio of 1.457, Table 5, model 2, Kennedy 2008). Micro-creativity was 30% more likely to occur with polite disagreements and 18% less likely to occur with rude disagreements (+30% = odds ratio of 0.772; -18% = odds ratio of -0.559; Table 5, Model 3). After a wrong idea in the previous turn (wrong idea [-1]), micro-creativity was 13% more likely than otherwise (odds ratio of 0.531, Table 5, Model 4, bottom). The final model accounted for 32% of the variance in micro-creativity (Table 5, Model 4, last row).

Table 5

*Summary of Standardized Regression Coefficients (and Their Standard Errors) in the Second SDA Results Modeling Micro-Creativity in Twenty Groups (n = 3,215 Turns of Talk)*

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variables added</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Summary time period</td>
<td>1.457***</td>
<td>1.393***</td>
<td>1.348***</td>
<td></td>
</tr>
<tr>
<td>(0.229)</td>
<td></td>
<td>(0.231)</td>
<td>(0.232)</td>
<td></td>
</tr>
<tr>
<td>Politely disagree</td>
<td>0.772***</td>
<td>0.702***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.123)</td>
<td></td>
<td>(0.125)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rudely disagree</td>
<td>-0.559*</td>
<td>-0.585*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.227)</td>
<td></td>
<td>(0.233)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Other current speaker variables Yes</td>
<td></td>
<td></td>
<td></td>
<td>0.531***</td>
</tr>
<tr>
<td>(0.152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrong idea (-1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Other previous speaker variables Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance at each level</td>
<td>0.177</td>
<td>0.188</td>
<td>0.174</td>
<td>0.193</td>
</tr>
<tr>
<td>Group</td>
<td>0.083</td>
<td>0.128</td>
<td>0.133</td>
<td>0.122</td>
</tr>
<tr>
<td>Time period</td>
<td>0.044</td>
<td>0.069</td>
<td>0.112</td>
<td>0.137</td>
</tr>
<tr>
<td>Turn of talk</td>
<td>0.158</td>
<td>0.215</td>
<td>0.276</td>
<td>0.315</td>
</tr>
</tbody>
</table>

Note. A constant term was included. * p < .05; ** p < .01; *** p < .001

This second SDA yields superior results compared to the first SDA and shows how attributes at different levels (time period, turn of talk) are related to a target process (micro-creativity). Contrasting the well-fitting turns with the smallest residuals versus the poorly-fitting turns with the largest residuals in another IDA can further inform our
theoretical model.

Discussion

Faced with little mixed methods research (MMR) on discourse (Bryman 2006, 2007), we proposed and showcased a new MMR approach to studying discourse, SIDA. Like most MMR methods, SIDA provides mutually supportive evidence, complements each other’s strengths, and addresses each other’s weaknesses. Unlike most MMR methods, SIDA’s cycle of complementary IDA<=>SDA components inform each other’s subsequent analyses. After discussing the benefits of the SIDA cycle, we consider its implications for MMR and discourse research in education.

Benefits of SIDA’s IDA<=>SDA cycle

SIDA’s complementary IDA<=>SDA cycles illustrate an iterative strategy for analyzing large data sets and demonstrate how qualitative and statistical analyses can mutually inform each other’s subsequent analyses (see Table 6). The first IDA selects and scrutinizes key phenomena within a small subset of contrasting cases (Roelle & Berthold, 2015; e.g., successful vs. unsuccessful group problem solving) to understand them in context. Then, the first IDA develops theory by specifying constructs, identifying patterns, and informing the designs of decision trees. By doing so, the first IDA defines variables, creates a database, and generates hypotheses for SDA.

Table 6

<table>
<thead>
<tr>
<th>IDA/SDA step</th>
<th>Benefit</th>
<th>Informs next analysis</th>
<th>SIDA general methods</th>
</tr>
</thead>
</table>
| 1st IDA      | • Develop theory  
                • Understand phenomena in few cases  
                • Develop constructs  
                • Identify patterns  
                • Inform design of decision trees | • Generates hypotheses  
                • Defines variables | Qualitative analysis |
| 1st SDA      | • Test hypotheses in few cases  
                • Identify Breakpoints  
                • Identify turns that fit model well  
                • Identify turns that fit model poorly | • Pinpoints critical Statistical analysis data for further explication | |
| 2nd IDA      | • Refine theory  
                • Understand breakpoints  
                • Elaborate mechanisms at well-fit turns  
                • Differentiate mechanisms at poorly-fit turns  
                • Develop new constructs  
                • Inform design of decision trees | • Generates hypotheses  
                • Defines variables | Qualitative analysis |
| 2nd SDA      | • Test hypotheses on all data  
                • Identify turns that fit model well  
                • Identify turns that fit model poorly | • Pinpoints critical Statistical analysis data for further explication | |
This part of the IDA <=> SDA process resembles mixed methods approaches that generate theory via inductive qualitative research and then test them via quantitative research (Mayoh & Onwuegbuzie 2015, p. 98). However, SIDA continues through further cycles of IDA and SDA, which shift the weight of the analyses between qualitative and quantitative poles. Each analysis feeds the following analysis with information and attention foci, toward creating a combined mixed methods approach.

The first IDA’s concepts and their relations inform creation and application of decision trees to categorize the small subsample’s data for the first SDA. The first SDA tests the first IDA’s hypotheses, identifies critical breakpoints, and pinpoints turns of talk that fit the model well versus poorly. The SDA’s breakpoints, well-fit turns, and poorly-fitting turns target specific data for the second IDA to examine in detail.

The second IDA then explores the first SDA’s identified breakpoints, elaborates the theoretical mechanisms at the first SDA’s well-fit turns, and differentiates the theoretical mechanisms at the first SDA’s poorly-fitting turns. Compared to the first cycle of IDA+SDA, the second IDA refines the theory to create a superior explanatory model by developing new constructs, defining new variables, and revising hypotheses.

The second SDA then tests the first IDA’s hypotheses on the entire data set and estimates each relationship’s generality (idiosyncratic, individual-specific, group-specific, activity-specific… universal). It identifies well-fit turns and poorly-fitting turns that a subsequent IDA can explore to improve further the theoretic mechanisms.

An IDA <=> SDA cycle (or sequence) can yield superior explanatory models, be more flexible, and be more efficient than a parallel IDA and SDA. Unlike the initial explanatory model in the first cycle of IDA+SDA, the explanatory model in the second cycle benefits from IDA of preliminary SDA results (breakpoints, well-fit turns and poorly-fitting turns). Detailed IDA analyses of these specific instances can refine coarse constructs (e.g., disagreement) and specify theoretical mechanisms to yield a superior explanatory model to test in the second SDA.

Also, SIDA is flexible and can increase efficiency. Researchers can flexibly start and stop at any point in the above multi-step IDA <=> SDA cycle. For example, a pilot study might use first IDA → first SDA. Or, researchers with clear hypotheses can test them first with a small subset of the data (first SDA→ second IDA) or the entire data set (second SDA→ second IDA). As each analysis informs each subsequent analysis, they inform decisions regarding whether additional expertise or analysis is helpful or needed. Furthermore, analytic results help target subsequent analyses to specific data to improve efficient use of intellectual resources (e.g., first SDA → second IDA).

Hence, the SIDA approach yields superior explanatory models, is flexible, informs subsequent analyses, and might serve as a pure mixed methods approach. An entire class of mixed method approaches can use iterative cycles in which qualitative and quantitative methods alternate, feeding one another information and attention foci.

**Implications**

This study has implications for mixed methods research, especially educational research. In this study, we contextualized SIDA within discourse studies in education; however, our approach can benefit any researcher considering mixed methods for studying discourse across different fields (e.g., political discourse, Parmelee & Perkins...
In one view of pure mixed methods, qualitative and quantitative parts (a) have equal status and (b) are balanced in the middle between the extremes of pure qualitative versus pure quantitative studies (Johnson et al., 2007). However, SIDA’s two components, IDA and SDA span the major paradigms, so we view SIDA’s cycles of IDA and SDA as a third paradigm that exists outside this continuum (Johnson et al. 2007, p. 129). While a researcher can use IDA in a qualitative-only study or use SDA in a quantitative-only study, SIDA allows researchers to work within and across quantitative and qualitative paradigms, yielding genuine integration of these methods (Bryman 2007) to foster superior explanatory models and theories.

SIDA’s integration of quantitative and qualitative methods helps meet the different demands of education policymakers/stakeholders and classroom teaching. Policymakers and stakeholders want quantitative measures of student performance and growth (performance capacity) to make simple, clear evaluations of students, teachers and schools (e.g., large-scale assessment studies with standardized tests [such as National Assessment of Education Progress] or experiments, Lagemann 2000; Pearson 2007; Phillips & Burbules 2000). Quantitative measures can capture performance outcomes and model organizational structures within positivist or postpositivist paradigms (e.g., peer-to-peer assistance [Crouch & Mazur 2001] or cooperative groups [Heller & Heller 1999]). However, quantitative measures lack the micro-level detail to reliably evaluate classroom discourse processes (e.g., ask questions, construct arguments, develop models, and explain content) specified in recent standards (e.g., Common Core State Standards, Next Generation Science Standards). Hence, researchers often use qualitative methods within social constructivist and discursive perspectives to examine how teachers can help their students engage in these desirable classroom processes (performance style, Gee 2014; Potter 2004). A transformative mixed methods paradigm such as SIDA addresses both performance capacity and performance style, enabling both (a) sufficient reliability, validity and generalizability to test hypotheses and apply ideas across education contexts, and (b) sufficient analytic detail to provide actionable guidance for teachers and students.

Limitations

MMR scholars face practical and epistemological challenges. Mixed methods research can take more time, require collaboration, and have fewer suitable venues for publication or dissemination. Although SIDA can save some time by identifying specific data points for IDA and specific elements to measure for SDA, conducting multiple qualitative and statistical analyses generally requires more time. Furthermore, as few researchers have expertise in both qualitative and statistical analyses (Jarosz et al., 2017), mixed methods research often requires collaboration. Furthermore, the word limits of many journals challenge mixed methods research scholars to present their multiple methods, analyses, and results within the same space as mono-method researchers. Without appropriate venues, the predilection for mixed methods research to represent one method more fully over the other will likely continue. Online venues might mitigate this constraint with additional space for extended appendices or longer articles at low cost (unlike print-only articles).

Conclusion

Confronted with education policymakers’/stakeholders’ demands (for reliability,
validity and generalizability) and teaching demands (for improving classroom discourse),
education researchers often address only one of these issues via a quantitative or
qualitative study. In this study, we propose and illustrate an approach that helps address
both sets of demands simultaneously, statistical interpretive discourse analysis (SIDA),
which combines cycles of interpretive discourse analysis (IDA) and statistical discourse
analysis (SDA) to yield superior theories and empirical results.

We illustrated how the first IDA on two contrasting cases (Roelle & Berthold,
2015) in a subsample develops theory by specifying constructs, identifying patterns, and
informing decision trees. These elements help create a database and hypotheses for the
first SDA to test. The first SDA supports or rejects these hypotheses, identifies critical
breakpoints, and pinpoints turns of talk that fit the model well versus poorly for the
second IDA to examine. The second IDA scrutinizes these turns to elaborate and
differentiate theoretical mechanisms, often developing new constructs and hypotheses
for the second SDA to test, and so on. Hence, mixed methods like SIDA can help
researchers address a wicked, complex problem in education—understanding classroom
discourse and its role in learning and teaching.

REFERENCES
Agar, M. (2015, Mar 3). Basic vs Applied vs. Practice, the End of Versus [Web Blog
Rates of Methodological Approaches across the Social/Behavioral Sciences.
peer video analysis events and literacy teachers’ outcomes. Journal of Technology
and Teacher Education, 23(2), 159-183.
Contexts. In M. F. Schober, D. N. Rapp & M. A. Britt (Eds.), The Routledge
Procedures that Control the False Discovery Rate. Biometrika, 93(3), 491-507.
doi: 10.1093/biomet/93.3.491
Analysis and the Study of Classroom Language and Literacy Events (1st ed.).
Publications.
Qualitative Research, 6(1), 97-113. doi: 10.1177/1468794106058877


Appendix A

Searching *Journal of Mixed Methods Research* and *Sociological Methods and Research* for the terms discourse analysis or conversation analysis

Using the JMMR and SMR websites we searched for the phrase discourse analysis and the more specific term conversation analysis (Schegloff, 2007). Searches for classroom discourse terms such as peer discussion or peer talk yielded no results. We cross-checked our search results with other databases that index JMMR and SMR (e.g., PsychINFO, SocINDEX). These searches yielded no additional articles.

**JMMR.** In total, 21 articles published in JMMR following Bryman’s (2006) extended review review have used the terms discourse analysis (DA) or conversation analysis (CA) in the full text, abstract, keyword or references. (See list at end of Appendix A). No JMMR articles used CA or DA in the title. CA appeared in only five articles whereas DA has appeared in 19 articles, and both DA and CA appeared in three articles. DA or CA appeared in the full text of 17 articles, and in four articles they appear only in the references or notes. When DA or CA appear in the full text, they are mentioned only 1-3 times and exclusively as qualitative research methods in 15 of these 17 articles. One article (Hayden & Chiu, 2015) examined reflective practices of preservice teachers using statistical discourse analysis and mentioned DA five times. Another article (Rinne and Fairweather, 2012) combined discourse analysis and cultural consensus analysis to explore New Zealand’s innovation success, mentioning DA 31 times.

**Journal of Mixed Methods Research Articles**


