From Moves to Sequences: Expanding the Unit of Analysis in the Study of Classroom Discourse

Adam Lefstein

Ben-Gurion University of the Negev

Julia Snell

University of Leeds

Mirit Israeli

Ben-Gurion University of the Negev

Note: The authors gratefully acknowledge support from the UK Economic and Social Research Council (*Towards Dialogue: A Linguistic Ethnographic Study of Classroom Interaction and Change*, RES-061-25-0363).

Abstract

What is the appropriate unit of analysis for the study of classroom discourse? One common analytic strategy employs individual discourse moves, which are coded, counted and used as indicators of the quality of classroom talk. In this article we question this practice, arguing that discourse moves are positioned within sequences that critically shape their meaning and effect. We illustrate this theoretical claim through exploration of a corpus of over 7000 discourse moves in primary literacy lessons. First, we use conventional measures such as the proportion of open and closed questions, and show how these indicators can be misleading when abstracted from the sequences in which they are embedded. We propose a complementary method, lag sequential analysis, which examines how discourse is sequentially structured – i.e. which discourse moves are followed by which other moves, and which chains of moves are more or less significant. We illustrate this method through re-analysis of our corpus of literacy lessons, examining differences between the sequential patterns found to be significant in the different classrooms observed. While lag sequential analysis does not resolve all problems inherent in systematic observation of classroom discourse, it does shed light on critical patterns in the data-set that would have otherwise gone unnoticed.

From Moves to Sequences: Expanding the Unit of Analysis in the Study of Classroom Discourse

Relating educational policies, classroom processes and pupil outcomes requires a reliable and valid means of measuring classroom discourse and interaction. A popular and expedient strategy involves systematic observation of discourse moves according to a predetermined coding scheme (typically focusing on actor and function) and statistical analyses of the resulting frequencies and correlations. Through such methods researchers have advanced understanding about relationships between the relative dialogicality of lessons and pupil achievement (Nystrand et al., 1997), about the extent to which new instructional policies have modified classroom interactional patterns (Galton et al., 1999; Smith et al., 2004), about the relationships between classroom interaction and social class (Harris & Williams, 2012), about shifts in classroom discourse following an extended collaborative action research project (Wells & Arauz, 2006), and about which instructional practices are associated with student explanations (Webb et al., 2009). In our own work, we have used such a strategy to identify from within a large corpus relatively dialogic episodes and to explore processes of continuity and change in classroom interactional patterns (Snell & Lefstein, 2011).

Systematic observation and coding of discourse moves poses, however, a number of problems. Mercer (2010), in a discussion contrasting different methods for analysing classroom discourse, raises three sets of issues. First are challenges with regard to coding reliability: meaning is inherently ambiguous, and is furthermore situated temporally (e.g. the same utterance will serve different functions at the beginning and at the end of a lesson). Second, validity problems: the relationship between linguistic form and discourse function is not as

straightforward as coding schemes might suggest (see also Gee, 2011, pp. 63-65; and Bloome et al., 2005, p. 32). Third, are "difficulties in determining the appropriate size of the unit of analysis to be coded... [f]or example, is the most meaningful unit a question or question-and-answer?" (p. 4).

This article addresses Mercer's question about the size of the unit of analysis. We propose that, for many purposes it is necessary to expand the unit of analysis beyond the individual discourse move in order to investigate how discreet moves are sequentially structured. In developing this argument, we first review the unit of analysis problem and its implications for the study of classroom discourse. Second, we discuss problems with the popular analytic strategy of employing the individual discourse move as unit of analysis, as illustrated in our analysis of 30 video-recorded literacy lessons in one London primary school. Third, we re-analyse the same data-set using lag sequential analysis (Bakeman & Gottman, 1997) to uncover sequential patterns within the corpus, and discuss advantages and limitations of this method.

Concerns with the Unit of Analysis

The current methodological concern with units of analysis has its roots in Gestalt psychology and, specifically, the way their ideas were developed and promoted by Vygotsky (Matusov, 2007). Briefly, Vygotsky argued against the traditional psychological method of decomposing complex mental processes into component elements that no longer captured the key characteristics of the whole. In a famous passage Vygotsky (1987) compared such a strategy to studying water through analyses of hydrogen and oxygen in isolation:

> When one approaches the problem of thinking and speech by decomposing it into its elements, one adopts the strategy of the man [person] who resorts to the decomposition of

> water into hydrogen and oxygen in his search for a scientific explanation of the characteristics of water, its capacity to extinguish fire or its conformity to Archimedes law for example. This man will discover, to his chagrin, that hydrogen burns and oxygen sustains combustion. He will never succeed in explaining the characteristics of the whole by analyzing the characteristics of its elements. (quoted in Matusov, 2007, p. 308)

Likewise, we will not be able to capture the distributed nature of cognition through study of individuals' thinking, isolated from the people and artifacts that mediate cognitive processes. Or, as we argue below, the co-constructed nature of discourse cannot be captured through study of individual discourse moves, isolated from the sequences of utterances in which they are embedded.

Instead of decomposing the object of inquiry into *elements*, Vygotsky suggests that researchers should "partition" the whole into *units* that preserve "all the basic characteristics of the whole". In this approach, the molecule is the smallest unit of analysis for studying H_2O, and word meaning (Vygotsky, 1987), mediated action (Wertsch, 1998) or activity systems (Engeström et al., 1999) are appropriate units of analysis for socio-cultural psychological research.

Matusov (2007), to whom this section is heavily indebted, discusses a range of methodological problems with overly reductive units of analysis. The most problematic is what he terms "horizontal reductionism", which "involves treating a part of a system as if it is the self-contained and isolated whole" (p. 315). For an example of horizontal reductionism consider an analysis of disability that treats the phenomenon as entirely a matter of individual capacity, without consideration of the social and material context against which differences in individual capacity are consequential, i.e. are *disabling* (see, e.g., Varenne & McDermott, 1998). For an

example from the area of classroom discourse consider an analysis of the functions of teacher questions that ignores the ways in which students respond to those questions. In both cases, an element of the system (individual capacity, a teacher question) is treated as if it operated independently of the system in which it is actually embedded.

The context-dependent nature of discourse is a well-developed theme in conversation analysis and linguistic anthropology (e.g. Duranti & Goodwin, 1992). The meanings and significance of utterances are shaped by a variety of contexts, including, for example, the cultural setting and social sphere; the activity in which interlocuters are engaged (Levinson, 1979); and even the actions of the audience while a speaker fashions their utterance (Schegloff, 2001). Conversation analysis emphasizes, in particular, the critical role of *sequential* context:

> No empirically occurring utterance ever occurs outside, or external to, some specific sequence. Whatever is said will be said in some sequential context, and its illocutionary force will be determined by reference to what it accomplishes in relation to some sequentially prior utterance or set of utterances. (Heritage & Atkinson, 1984, p. 6)

In light of this principle, conversation analysts attend closely to the sequential unfolding of discourse, to how each turn addresses that which preceded it, and to the range of possible responses it projects for the next speaker. Hence, for conversation analysts, the key unit of analysis (though they don't explicitly discuss methodology in such terms) is the exchange, and the minimal unit is an adjacency pair, i.e. two turns that conventionally come together, with the first turn setting up the expectation of the second, such as question and answer, apology and minimization, or invitation and acceptance.[1]

[1] Note that conversation analysts tend to use the term "unit" differently than I am employing it here. Hence, for example, when Selting (2000) refers to the Turn Constructional Unit (TCU) as "the basic unit of talk suggested by Conversation Analysis" (p. 477), she means that it is "the smallest possible complete linguistic [unit] in [its] context

Sociocultural researchers seek to replace reductive units of analysis with more holistic ones. However, holistic units are not without their pitfalls, chief among which are issues of manageability. Moreover, it seems that no unit of analysis can ever be big enough, as all social practices are embedded in larger social, economic and cultural systems. "Each new candidate for the unit of analysis," remarks Matusov, "...sooner or later becomes recognized by sociocultural scholars as 'too small'. A unit that preserves the whole of the phenomenon, as Vygotsky and Gestalt psychologists insisted, seems to resist having its boundary and limit" (p. 323). Discourse analysis similarly resists boundaries, and often analysts find that interpretation of an utterance requires consideration of not only the social situation and activity, but also the entire history of the discourses spoken. As Bakhtin (1981) wrote, "each word tastes of the context and contexts in which it has lived its socially charged life" (p. 293); all of these contexts are potentially relevant to understanding the use of that word.

In attempting to strike a balance between reductionism and holism, we will argue below that interactional structures of three or more turns is a methodologically expedient and theoretically sound unit for the analysis of classroom discourse (at least for many common purposes and contexts).

Matusov notes that researchers often choose their unit of analysis in part because of institutional constraints or momentum, i.e. "because the research method for which this unit of analysis will be used is well developed and not because it is the best fit for the study of the conceptualized construct of the phenomenon" (p. 314). In such a way, for example, we adopted methods for systematic observation, coding and counting frequencies of discourse moves for our study of continuity and change in classroom discourse. While these methods were in retrospect

---

of interaction" (p. 512) rather than the smallest unit that preserves all the basic characteristics of the whole (as per Vygotsky).

not the best fit for our purposes (to be discussed below), we should note that it makes a lot of sense for researchers to use existing and accepted methods, in order to benefit from others' investments in developing, trialling and refining them, and in order to afford between-study comparisons. We turn to our own research in the next section.

Discourse Move as Unit of Analysis in the *Towards Dialogue* Project

The impetus for our work on this topic comes from our experiences with systematic observation of classroom discourse data in the *Towards Dialogue: A Linguistic Ethnographic Study of Classroom Interaction and Change* project.[2] This study examined processes of continuity and change in classroom interactional patterns through quantitative interrogation of systematic observation data (the focus of this article) and linguistic ethnographic micro-analysis of select segments. Details about this project are elaborated elsewhere (Lefstein & Snell, 2011a, 2011b, 2011c, 2013, 2014; Snell & Lefstein, 2011); here we briefly describe details particularly relevant to the current article.

We observed and video-recorded naturally occurring classroom discourse and interaction in 73 literacy lessons in seven upper primary classrooms at Abbeyford Primary School[3], a relatively large community primary school in an East London borough with a long-standing interest in dialogic pedagogy and a history of developing and implementing pedagogical innovations. A senior local authority advisor recommended Abbeyford Primary on account of its highly regarded, stable and experienced teaching staff and leadership team. We also conducted a bi-weekly professional development workshop for participating teachers, in which we assisted in

[2] We gratefully acknowledge support received from the UK Economic and Social Research Council (ESRC), award no. RES-061-25-0363.
[3] A pseudonym, as are all the proper names in this article.

lesson planning, guided reflection on video-recorded classroom episodes and discussed teachers' concerns in the enactment of dialogic pedagogy.

In order to investigate continuity and change in classroom interactional patterns we subjected a sub-set of 30 lessons to computer-assisted systematic classroom observation, sampled from lessons taught by three teachers, who had between 10 and 11 years of teaching experience and who had also been involved in a previous dialogic teaching programme. Analyses of this coded corpus were used to situate the sample (by comparing it to previous studies), to examine processes of change over time, to investigate correlations between relative dialogicality and other key variables (especially teacher and pedagogic activity), and to select episodes for micro-analysis.

Systematic observation of these lessons focused only on the whole-class teaching segments (defined as a whole class activity lasting longer than 2 minutes). This accounted for approximately 50 percent of the total duration of the lessons (i.e. 24 minutes of an average 48 minute lesson). For each whole-class segment we coded discourse moves using the systematic observation software, *The Observer XT*[4] (Noldus 2008), using a coding system adapted from that developed by Hardman and colleagues (Hardman et al. 2003; Smith & Hardman 2004). This system codes discourse moves, defined as a single utterance or a string of uninterrupted utterances with a common function (e.g. to explain, direct, question, respond, give feedback). Questions were further subdivided according to type (e.g. 'open', 'closed', 'probe'). We also distinguished between 'simple' feedback (repetition of a pupil answer or very brief response such as 'Okay') and 'elaborated' feedback (an extended response). We further coded for activity type (e.g. 'Recap', 'Review of group work', 'Introduce new task'). In the way we set up the system, the software "stopped" one move once another move was coded. Noticeable pauses

[4] For a review of this software see Snell (2011).

between moves were coded as "silence". Because the start of every code is time-stamped, durations as well as frequencies of coded behaviours were recorded.[5]

Analysis of this corpus yielded a number of interesting yet tentative findings related to (a) similarities and differences between teachers, and between the classroom practice observed and a national sample of literacy lessons; (b) changes over time in classroom interactional patterns; and (c) the conditions under which teaching and learning was more or less dialogic. We also used the analysis to locate particularly dialogic episodes – i.e. episodes with a large proportion of open, probe or uptake questions, elaborated feedback, and high rates of pupil participation – which we then subjected to further, micro-analytic scrutiny.

We briefly outline and illustrate some key findings, and then discuss their limitations, especially with regard to the unit of analysis. In the interests of brevity we focus in particular on teacher question types. The use of teacher questions as a key indicator of discourse quality is relatively common in such research, with a high proportion of closed, factual, or exam questions interpreted as evidence of low cognitive demand, and conversely, a relatively high proportion of open, uptake or probe questions interpreted as a sign of productive dialogue (e.g. Galton et al., 1980, 1999; Goodlad, 1984; Myhill & Dunkin, 2005; Nystrand et al., 1997; but see also Dillon, 1982, for a dissenting view).[6]

Contrasting Classroom Discourse: Between Teachers and with a National Sample

[5] Further details on the coding system, including definitions for all the codes, can be found in Snell & Lefstein (2011).

[6] We used the following operational definitions for the central question types:
Open question: a question for which the teacher does not appear to have a prespecified answer in mind.
Closed question: a question for which there are a limited range of prespecified acceptable answers.
Probe question: a follow-up question designed to extend an individual pupil's response.
Uptake question: a follow-up question in which the teacher incorporates a pupil's answer into a subsequent question directed to the whole class.

Table 1 details average rates and durations of discourse moves in each of the classrooms, and contrasts these data with findings from a national study of literacy lessons collected by researchers at the University of Newcastle in 2001 (Hardman et al. 2003). The numbers in the top half of the table show the 'rate' (i.e. number per hour) for teacher and pupil discourse moves. Rate is calculated as frequency per hour to make this data comparable to other studies. If, for example, a teacher used 5 open questions in 20 minutes of whole-class teaching, this would be reported as a rate per hour of 15. Rate is recorded for each individual teacher and for the school as a whole (i.e. the average for all 3 teachers), and this is compared with the averages reported by Hardman and colleagues for the 35 literacy lessons included in their national sample (Hardman et al. 2003; Smith et al. 2004).[7]

Insert Table 1 approximately here

The table shows that teachers at Abbeyford Primary asker fewer closed questions than the teachers in the national sample, and instead adopted a more dialogic stance, using more open questions (i.e. questions for which there is no single, predefined correct answer) and probes (where the teacher stays with the same pupil to extend their initial response). The percentages to the left of 'rate' show each question type as a percentage of total questions posed. While 50% of

[7] Note that there is not a one-to-one correlation between the categories adopted in our analysis and those used in the Newcastle study. For example, results from the Newcastle study did not differentiate between elaborated and simple feedback. Further, a number of teacher discourse moves in the Abbeyford Primary data were coded as 'response to pupil', a category not present in the analysis of the national sample. This category includes responses to pupil questions, but it also incorporates discourse moves which did not neatly fit into other categories (e.g. statements which were neither 'explain' nor 'feedback'), and which tended to fall outside of the canonical *Initiation-Response-Evaluation* (IRE) cycle.

questions in the national sample were closed, only 34% of questions at Abbeyford Primary were of this type.[8]

Likewise, the analysis suggests that alongside the differences between the school and the national sample there are also meaningful differences between teachers and classrooms within the school. In particular, Ms Leigh's classroom stands out as exhibiting much higher rates of elaborated feedback, and both she and (to a lesser extent) Mr Robbins posed a greater proportion of open and probe questions than Ms James.

Changes over Time in Classroom Interactional Patterns

Systematic observation of discourse also allows us to measure the extent to which classroom discourse has changed over time. For example, Figure 1 tracks the ratio of open to closed questions in the thirty observed lessons (aggregated for all three teachers by order of observations), showing that over the course of the study, the teachers began to ask more open questions and fewer closed questions.

Insert Figure 1 approximately here

This movement towards more frequent use of open questions and away from closed questions suggests a change in teaching practice towards more dialogic pedagogy. However, this was the only appreciable change over time, which raises the concern that this development was merely a 'bolt-on' to familiar practice (Galton et al., 1999: 52) rather than an indicator of more penetrating and/or durable changes in teaching and learning.

[8] The proportion of open questions in both sets of data represents a considerable increase on the findings of the earlier ORACLE 1976 study, where open questions formed only 5% of all questioning (Galton et al. 1980: 87), and of the follow-up study in which 12.8% of questions asked in English lessons were open (Galton et al. 1999: 74).

Conditions under which Teaching and Learning Was More or Less Dialogic

We looked at a number of different factors shaping fluctuations in the frequency of moves that are perceived to be indicative of relatively dialogic discourse; here we focus on the distribution of discourse moves between pedagogic activities. Discourse features indicative of relatively dialogic teaching clustered around certain pedagogic activities, such as review of group work, text-based discussions and feedback on pupil writing, and this phenomenon was particularly pronounced in one of the classrooms (see figure 2). When the teacher, Ms. Leigh, introduced a new task she posed more closed than open or probe questions (55% closed vs. 41% open and probe questions), but this ratio was inverted when discussing texts (15% closed vs. 83% open and probe questions), reviewing pupil writing (19% closed vs. 71% open and probe questions) or engaging in role play activities (5% closed vs. 80% open and probe questions).[9]

Insert Figure 2 approximately here

Uneasiness with the Move as Unit of Analysis

The above interpretations of our findings are based on the assumption that teacher questions are a good indicator of the dialogic quality of the discourse overall. Though this assumption is generally accepted, or at least implicit in most relevant studies, we were uneasy with the reductionism involved in such an extrapolation from discourse move (teacher question) to interactional category (dialogic pedagogy). First, the category of dialogic pedagogy includes multiple dimensions of discourse and activity, among them participation norms, interpersonal

[9] Detailed analysis of these patterns, including break-down of frequencies of all discourse features by activity type for each classroom, is provided in Snell & Lefstein (2011).

relations, substantive content and epistemological stance (see Lefstein, 2010; Lefstein & Snell, 2011, 2014). The ratio of open to closed questions is only relevant to the final dimension: closed questions are assumed to be suggestive of an authoritarian epistemological stance, and vice-versa. But, this assumption is also problematic: the educative qualities of dialogic interaction do not derive in and of themselves from teacher questions, but rather from the subsequent student participation and teacher follow-up that are assumed to be stimulated by such questions (see e.g. Cullen, 2002; Nassaji & Wells, 2000). However, the common idea that open questions will lead to and be associated with other (roughly, dialogic) discourse features is undermined by the finding (noted above) that no other aspects of discourse changed alongside the change in the ratio of open to closed questions. The fact that only question types changed should make us suspicious of the idea that that shift reflects a change in the quality of interaction.[10] Finally, when we looked more closely at 19 episodes that boasted a high density of "dialogic" discourse moves (open, uptake and probe questions; elaborated feedback; pupil response to pupil; and relative pupil participation), seven of them came up as "false positives", i.e. episodes that did not meaningfully differ from the rest of the corpus with regard to the substantive issues that interested us, such as the interplay of teacher and student voices, reciprocal discourse norms, and a critical, exploratory approach to knowledge.

To illustrate the importance of expanding the unit of analysis from discourse move to sequence, consider the following extract, which took place at the beginning of a Year 6 lesson on waste and recycling. The teacher, Ms. James, projected a powerpoint slide that posed an apparently open question: "What do we do with our rubbish?"

[10] See Lefstein (2008) for an account of how pupils can "close" down a series of teacher open questions.

Extract 1. "All that rubbish"[11]

1	Ms James:	right
2		well what do we do with our rubbish (1.5)
3		what does happen to our rubbish
4		have a think
5		about what happens to your rubbish
6		that you get rid of at home
7		do you actually think about
8		what happens to your rubbish
9	pupils:	[no:
10	Ms James:	[do you just
11		use something and then
12		just put it in the bin
13		and just (xxxxx xxxx)
14		it's in the bin
15		nothing to do with me anymore
16	pupils:	xxx xxxx *((a few shake heads))*
17	Ms James:	where does it actually go
18		what do we actually do with it
19		Bethany
20	Bethany:	it goes to landfill
21	Ms James:	right landfill
22		we mentioned this the other day
23		what is landfill
24		what is landfill
25		any ideas
26		Nigel what's landfill
27	Nigel:	(putting) the rubbish in the ground
28	Ms James:	yeah
29		it's the rubbish that goes in the ground

A quick tally of discourse moves shows that this 30 second segment features a high ratio of teacher open to closed questions. Ms James poses a series of three open questions (lines 2, 21-22, and 24-29) and only one closed question (line 23). However, though she explicitly instructs the pupils to "have a think", the open questions do not develop into a particularly thoughtful or productive exchange. Instead, after posing and repeating her first question (what

[11] Transcription notations:
(text) - Transcription uncertainty
xxxx - indecipherable speech
(1.5) - A 1.5 second pause
(( )) - Description of non-verbal activity
[- Overlapping talk or action
[
text - Emphasised relative to surrounding talk (underlined words)
te:xt - Stretched sounds

happens to our rubbish?), Ms. James poses a second question (do you even think about this question?), reformulates it in more concrete terms (do you just put it in the bin… nothing to do with me anymore?) and then returns to a slightly different version of her first question (where does it actually go?). At this point she receives and acknowledges a pupil response (line 20), which she then uses as the basis for a classic closed elicitation sequence (Mehan, 1979) on lines 22-29. Clearly, counting the frequencies of open and closed questions – without attending to the ensuing interaction – paints a rather misleading picture of the academic quality of the talk in the segment. In the next section we discuss alternative methods of analysing classroom discourse that attempt to account for its sequential development.

Expanding the Unit of Analysis: Sequential Analysis

Current attempts to account for the sequential nature of classroom interaction can be divided into top-down and bottom-up methods.

Top-down methods seek to code units larger than the discourse move, such as genre, activity or exchange. Sinclair & Coulthard (1976), for example, in their seminal study of the structure of classroom discourse, developed a hierarchical coding scheme in which lessons are composed of transactions, which contain teaching exchanges, which comprise acts (discourse moves). Their analysis highlighted the prevalent Initiation-Response-Follow-up (IRF) structure, which has formed the basis for much classroom discourse analysis since, including the coding system we have adapted here. Building on Sinclair and Coulthard's scheme, Wells and colleagues (Wells, 1999; Wells & Arauz, 2006) coded episode activity orientations and types of exchanges, in addition to discourse moves. Because of the limited size of the corpus, and the large variability among episode durations, their quantitative analyses focused on the discourse

moves within the episodes, rather than on the episodes themselves. For example, they investigated frequencies of negotiatory and known information questions within the various episodes (and learning domains), the lengths of student responses, and the extent to which responses to the different types of questions were followed up with evaluative or other forms of teacher feedback.

Another example of coding discourse units larger than the discourse move is Mercer and colleagues' (Mercer, 1996; Mercer & Littleton, 2007) distinction between exploratory, disputational and cumulative forms of talk. These categories have been productively used to guide practice, and trained observers have been able to use them to differentiate between experimental and control groups, but, as Mercer (1996) explains, they "are not meant to be descriptive categories into which all observed speech can be neatly and separately coded (as might be done in systematic observation research)" (p. 369). Rather, when Mercer and colleagues use quantitative measures of discourse in their research they turn to frequencies of key phrases such as "I think" and lengths of utterances (Mercer & Littleton, 2007; Mercer, Wegerif & Dawes, 1999).

Likewise, in most of our own work, we have used micro-analytic methods to examine closely episodes lasting several minutes or longer (e.g. Lefstein & Snell, 2011c, 2014). Such methods involve working slowly through the fine-grain details of talk and interaction, moving between the specific event and the broader social structures, institutions, discourses and histories that shape it, and of course consulting with the participants themselves. This strategy, however, is extremely time-consuming and, while analysing episodes or genres makes good sense theoretically, these categories tend to be relatively loose, with fuzzy boundaries, and as such are difficult to operationalize and study quantitatively.

An alternative approach is to work bottom-up from discourse moves to examination of interactional sequences through lag sequential analysis (Bakeman & Gottman, 1997; see also Chiu & Khoo, 2005, for a discussion that contrasts lag sequential with competing techniques). This statistical method facilitates interrogation of the sequential organization of a series of events (for example, discourse moves), including the conditional probabilities that some events will be followed by certain others, the strength of these associations, and the significance of the sequential patterns observed (i.e. that the relationships are not the result of the random distribution of independent events).

Two recent studies employ lag sequential analysis to study classroom discourse patterns. Jadallah and her colleagues (2011) coded small group discourse for different types of teacher scaffolding moves and pupil responses in a corpus of 30 Collaborative Reasoning discussions (including 5,300 turns at talk). They then analysed the probabilities of relevant pupil moves following teacher prompts (e.g. pupil clarification following teacher request for clarification), demonstrating both the immediate and delayed effects of teacher scaffolding moves on pupil responses, and of pupil responses on other pupils' utterances. Similarly, Molinari and colleagues (2012) used lag sequential analytic methods to investigate naturally occurring whole class teaching in three Italian primary schools. The coded teacher questions, pupil responses and teacher follow-up moves in a corpus of 828 triads identified in close to 10 hours of recordings, and examined conditional probabilities between teacher initiations (e.g. authentic vs. focused questions), pupil responses (e.g. correct vs. erroneous answers) and teacher follow-up moves. They found a variety of significant triadic chains, including for example teacher authentic question-correct pupil answer-teacher simple follow-up and teacher focused question-incorrect pupil answer-teacher scaffolding. We build on these techniques in this article.

Lag Sequential Analysis of the Toward Dialogue Corpus

We translated our corpus into the Sequential Data Interchange Standard (SDIS) and analysed it by means of the Generalized Sequential Querier (GSEQ) 5.1 and ILOG 4 computer programmes (Bakeman & Quera, 2011; Bakeman & Robinson, 1994). The resulting data-base includes 7,228 discourse moves (hereafter, *events*) divided into 142 discrete observations (a topically focused pedagogic activity in the whole class teaching segment in one of twenty-nine lessons taught by three teachers). The original corpus was coded using 25 discreet categories. However, performing lag sequential analysis on this number of codes would require many more events than we recorded (the formula for computing the necessary number of events is $5 \times K^{l+1}$ where K is the number of codes and *l* is the number of lags examined (lag +1 are two move chains, lag +2 yield three move chains, etc.); therefore, analysing 25 codes at lag +2 would require 78,125 events). For this reason, we selected the seven codes that most interest us theoretically – teacher open and closed questions, pupil brief, moderate and lengthy responses, and teacher simple and elaborate feedback – lumping all the other codes into an Other category.[12] Performing lag sequential analysis at lag +2 with eight codes necessitates 2,560 events, a requirement we satisfy for one of the teachers (Ms James) and of course for all of the teachers pooled together, but not for the different types of pedagogic activity; performing this analysis at lag +1 requires 320 events, a requirement we satisfy for all teachers and most of the activities. Counts and relative frequencies of the eight codes are provided in Table 2:

---

[12] We also simplified the original coding scheme by combining male and female pupils, and by removing the distinction between negative and positive feedback. We further recoded the mass of pupil responses to teacher (n= 2,180) in order to differentiate between brief responses of less than 1.42 seconds (PRB, n=549), moderate responses of between 1.42-5.69 seconds (PRM, n=1,084), and long responses of 5.7 seconds or longer (PRL, n=547).

Insert Table 2 approximately here

We posed five questions for lag sequential analysis of this data – the first two refer to general sequential tendencies within the corpus, the latter three to the specific issues explored above through examination of frequencies of individual discourse moves:

1) Is the observed discourse sequentially structured? Put another way, what is the likelihood that the structures observed are random, the result of the independent distribution of discourse over time?

2) Do certain kinds of pupil response follow certain kinds of teacher questions? And are these question-response pairs followed by certain types of follow-up? Specifically, we hypothesize (a) that Teacher Closed Questions are likely to be followed by Pupil Brief Responses, which in turn are likely to give rise to Teacher Simple Feedback (i.e. the chain TQC-PRB-TFS is hypothesized to recur significantly more often than if discourse moves were independent of one another), and (b) that Teacher Open Questions are likely to be followed by Pupil Long Responses, which in turn are likely to give rise to Teacher Elaborate Feedback (i.e. the chain TQO-PRL-TFE is hypothesized to recur significantly more often than if discourse moves were independent of one another).

3) Are there significant differences between the teachers with regard to how discourse in their classrooms is sequentially structured? For example, in the comparison of frequencies of key discourse moves Ms Leigh's class appeared to be more dialogic than Ms James' class – are these differences manifest also in sequential patterns?

4) Did the sequential patterns observed in the three classrooms change over time? For example, were teachers at Abbeyford Primary School more likely to use certain (dialogic)

discourse chains as they deepened their understandings of classroom discourse through collaboration with one another and with us in video-based reflective workshops?

5) Did the sequential patterns observed vary between different pedagogical activities? For example, certain activities, such as discuss texts and feedback on pupil writing, boasted higher frequencies of dialogic discourse moves – are there similar differences in sequential patterns?

We begin with the latter three questions, since we must first establish how stable the sequential dependencies are in order to decide how to cut up the data-set in order to investigate the first two questions. In the accepted terms of lag sequential analysis we need to test for *heterogeneity* – the stability of the sequential relationships across the three different teachers' classrooms, and *stationarity* – the stability of the sequential relationships over time (in our case, across pedagogic activities and over the course of the school year). To do so, we compared the sequential dependencies in sub-sets of the corpus (e.g. each of the classrooms) to the corpus as a whole and computed likelihood ratio chi square statistics to test for the significance of the differences between them (see, e.g., Jadallah et al., 2011). We based these tests on 2 X 4 contingency tables that include as the given event the two categories of teacher questions that most interest us (TQO and TQC), with the three types of pupil responses (PRB, PRM and PRL) as the target events. The results of these and the other tests are displayed in Table 3.

Insert Table 3 approximately here

The results of the heterogeneity tests are significant, indicating that the differences between the classrooms are greater than might be expected by chance.[13] For this reason, we do not pool the data across classrooms for the rest of the tests, instead analysing each classroom corpus separately.

We conducted two stationarity tests: one for changes over time, for which each classroom corpus was divided into three segments, and the other for differences between pedagogic activity (observations were pooled for each of eight pedagogic activities; one activity, Explain, was removed since it occurred only once in the corpus and contained only 62 events). Like the tests of heterogeneity, the stationarity tests are based on the sequential dependencies between closed and open teacher questions and brief, medium and long pupil responses.

Readers will note that degrees of freedom vary among the stationarity tests. These differences are due to the absence of the relevant sequential chains in some of the activities and stages. In such cases we either removed the relevant code (PRB or PRL) and/or activities, depending on which omission maximized the number of events in the test. The tests examine non-stationarity, i.e. the assumption that differences between the segments or activities are significant. Hence, a non-significant result shows that differences between the segments and activities are no greater than might be reasonably expected by chance, and we should therefore reject the hypothesis that the relevant discourse patterns changed over time or between pedagogic activities. The results of these tests are mixed: seven indicating stationarity and five demonstrating significant differences between activities or over time. The significant results of the three tests for stationarity between activities can be explained in each case by one particularly

[13] It is important to note that we have conducted a large number of significance tests (20 in table 3 alone), so we need to exercise caution in interpreting results. Consequently, we do not consider 0.05 p values to be significant, and below (table 4 and figures 3 and 4) only display as significant results that meet the 0.01 and 0.001 *p* values (using two and three asterisks respectively).

deviant pedagogic activity: review group work vis-à-vis TQC-initiated dyads in Mr Robbins' classroom, discuss texts vis-à-vis TQO-initiated dyads in Ms Leigh' classroom, and role play vis-à-vis TQO-initiated dyads in Mr Robbins' classroom. In all three cases, the results are insignificant if this one anomalous activity is removed. The significant results for the tests of change over time are due to a sharp reduction in TQC-PRM and TQC-PRL dyads in the final segment in Mr Robbins' classroom (conditional probabilities of 0.13 and 0.07 respectively, compared to 0.4 and 0.16 overall), and to a relative decrease in TQO-PRB and TQO-PRM dyads in the middle segment in Ms James classroom (conditional probabilities of 0.14 and 0.14 respectively, compared to 0.29 and 0.33 overall), likely due to intensive preparations for the standardized tests that took place in May of the academic year (see Snell & Lefstein, 2011, p. 13). On account of these significant results we focus our analyses primarily on Ms James' and Ms Leigh's classrooms, which exhibited fewer stationarity problems.

We now turn to examine the nature of the sequential patterns within each of the classrooms. First, we conducted tests of *independence*, i.e. the hypothesis that the discourse moves are independent of one another, and any relationships detected are the result of their random distribution. These tests are all significant at the 0.001 level, indicating that the sequential dependencies are highly unlikely to have occurred by chance. Next we explored specific chains of discourse moves that interested us: TQC-PRB-TFS and TQO-PRL-TFE. The first chain is typical of recitation teaching: the teacher poses a closed question, which is met by a brief (under 1.5 second) response, which is then followed up by simple feedback (acknowledgement or rejection of the answer without further explanation or elaboration). The second chain is indicative of more dialogic teaching: the teacher poses an open question, which is met by a long (over 5.7 seconds) pupil response, which is then followed up with elaborated

teacher feedback. Table 4 displays the distribution of these two chains within each classroom, along with *z* scores, which are used to calculate the significance of the chain given the baseline frequencies of each of the events within the relevant corpus.

Insert Table 4 approximately here

Another way of examining these relationships is by displaying graphically the associations between each of the phases. Figures 3 and 4 show the two most significant chains observed: TQC-PRB-TFS in Ms James classroom and TQO-PRL-TFE in Ms Leigh's classroom, juxtaposed alongside the sequential dependencies for other pupil responses and teacher feedback moves for the given teacher initiations. Note that these figures are not conventional state transition diagrammes, in which conditional probabilities are given for the relationships between contiguous events; rather, the diagrammes are cumulative: for lag +2 (i.e. two moves after the teacher initiation) we show the probabilities of the target follow-up move given the teacher question-pupil response dyad, and at lag +3 we show the probability of cycling back to TQC following the entire TQC-PRB-TFS chain (in figure 3). For each significant relationship we display conditional probabilities (written "p (I | J)", i.e. the conditional probability of I given J), Yule's *Q* (a value that ranges from 0 to 1, with 0 signifying no association, and 0.2, 0.43 and 0.6 conventionally interpreted as weak, moderate and large associations respectively; see Yoder & Symons, 2010, p. 126) and joint frequencies. Dotted arrows show non-significant relationships, while heavy lines show strong associations.

Insert Figures 3 and 4 approximately here

So as not to overwhelm, we have spared readers detailed information about the sequential dependencies of TQC and TQO chains in the other two classrooms, in which the associations are not nearly as strong or clear-cut. By way of illustration, the conditional probability of PRB following TQC in Ms Leigh's classroom is 0.19 (Yule's $Q = 0.56$), and the conditional probability of PRL following TQO in Ms James' classroom is 0.08 (Yule's $Q = 0.38$). The sequential analysis shows that the discourse norms that bind teacher closed questions, brief pupil responses and teacher simple feedback are particularly strong in Ms James' classroom, while the open question, extended pupil response and elaborate feedback pattern is most pronounced in Ms Leigh's classroom.

Conclusion: the Potential and Limitations of Sequential Classroom Discourse Analysis

This article has explored the problem of the unit of analysis in studies of classroom discourse. We noted that most attempts to quantitatively study classroom discourse use the individual discourse move as their unit of analysis. We questioned this practice, arguing that discourse moves are positioned within sequences that critically shape their meaning and effect. We illustrated this theoretical point through exploration of discourse data from a study of primary literacy classrooms, showing that accepted indicators such as open and closed questions can be very problematic if abstracted from the sequences in which they are embedded. We then proposed that lag sequential analysis may help overcome some of the shortcomings associated with conventional calculations of frequencies and rates of individual moves, and demonstrated this method in a re-analysis of the same data set. To what extent has the sequential analysis changed our understanding of the discourse analysed? What are its advantages and limitations?

The lag sequential analysis findings coincide to a certain extent with the comparison of rates of open and closed questions in Table 1. Ms James posed a greater proportion of closed to open questions than did Ms Leigh, and offered less elaborate feedback. Likewise, pupils in the former classroom speak for less time than in the latter classroom. However, the sequential analysis shows how these different discourse moves interact. After all, Ms James also posed open questions, and Ms Leigh closed questions. Significantly, the closed questions in Ms James' class tended to give rise to brief pupil responses, which in turn tended to receive simple teacher feedback followed by a further closed question: a statistically significant 71 occurrences of this chain in Ms James' classroom in comparison to a non-significant 5 instances in Ms Leigh's classroom (see table 4). Likewise, almost half of Ms Leigh's open questions led to lengthy pupil responses, which were strongly associated with elaborate feedback (a significant 6 occurrences of this chain in Ms Leigh's classroom compared to its non-signficant non-occurrence in Ms James' classroom). Consequently, we would argue that the meaning of open and closed questions in the two classrooms is qualitatively different, and that counts of frequencies of individual discourse moves do not tell the whole story.

Similarly, the sequential analysis offers some insight into the issues of change over time and differences among the pedagogic activities. Many of the differences that we observed in the distribution of frequencies (figures 1 and 2) are not accompanied by similar differences in sequential structures. While some of the non-stationarity tests for differences between pedagogic activities were significant, these effects were not nearly as pronounced as in the comparison of individual discourse move frequencies, and disappeared when one anomalous activity was omitted. Finally, only one of the two significant non-stationarity tests for changes over time – that of TQC-initiated sequences in Mr Robbins' class – supports the finding that discourse

became more dialogic over the course of the year. Moreover, this finding is not supported by the stationarity test for TQO-initiated sequences in that classroom, or for the tests of stability over time in the other two classrooms.

In closing, we note some key limitations of this exercise. First, we did not have sufficient data to perform some of the tests we would have liked to have undertaken, for example, comparing lag +2 sequential patterns in the different pedagogic activities. Second, our original coding system was not devised for this sequential analysis, and as a result included too many codes and, critically, insufficiently sensitive categories to differentiate pupil responses. In future research we intend to use coding schemes that differentiate more clearly between types of pupil discourse moves, investigating, for example, the accuracy of pupil answers, the use of argumentation, and speaking with or without explicitly receiving the floor. Third, and most importantly, lag sequential analysis cannot overcome the inherent limitations of a coding system that focuses on form rather than meaning or relationships, and is incapable of coming to terms with ambiguity.

These limitations notwithstanding, we believe that this exercise has demonstrated the potential power of sequential analysis to address some of the shortcomings of conventional methods for counting relative frequencies and distributions of individual discourse moves. Units of analysis of course need to be appropriate to research purposes, but they also need to be well-suited to the intrinsic qualities of the object of inquiry. Lag sequential analysis offers a promising way forward in accounting for the sequentially structured nature of classroom discourse, in a manageable and quantifiable manner.

References cited

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: an introduction to sequential analysis* (2nd ed.). Cambridge: Cambridge University Press.

Bakeman, R., & Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Bakeman, R., & Robinson, B. F. (1994). *Understanding log-linear analysis with ILOG: An interactive approach*. Hillsdale, NJ: Erlbaum.

Bakhtin, M. M., Holquist, M., & Emerson, C. (1981). The dialogic imagination: four essays. Austin: University of Texas Press.

Bloome, D., Carter, S. P., Christian, B. M., Otto, S., & Shuart-Faris, N. (2005). *Discourse analysis & the study of classroom language & literacy events: a microethnographic perspective*. Mahwah, N.J.: L. Erlbaum Associates.

Chiu, M.M., & Khoo, L. (2005). A new method for analyzing sequential processes dynamic multilevel analysis. *Small Group Research, 36*(5), 600-631.

Cullen, R. (2002). Supportive teacher talk: The importance of the F-move. *ELT Journal, 56*(2), 117-127.

Dillon, J. T. (1982). The Effect of Questions in Education and Other Enterprises. *Journal of Curriculum Studies, 14*(2), 127-152.

Duranti, A., & Goodwin, C. (1992). *Rethinking context: language as an interactive phenomenon*. Cambridge: Cambridge University Press.

Engeström, Y., Miettinen, R., & Punamäki, R.L. (1999). *Perspectives on activity theory*. Cambridge, UK: Cambridge University Press.

Galton, M. J., Croll, P., & Simon, B. (1980). *Inside the primary classroom*. London ; Boston: Routledge & Kegan Paul.

Galton, M., Hargreaves, L., Comber, C., Wall, D., & Pell, T. (1999). Changes in Patterns of Teacher Interaction in Primary Classrooms: 1976-96. *British Educational Research Journal, 25*(1), 23-37.

Gee, J. P. (2011). *An introduction to discourse analysis: theory and method* (3rd ed.). Milton Park, Abingdon: Routledge.

Goodlad, J. I. (1984). *A place called school: prospects for the future*. New York: McGraw-Hill Book Co.

Hardman, F., Smith, F., & Wall, K. (2003). 'Interactive whole class teaching' in the national literacy strategy. *Cambridge Journal of Education, 33*(2), 197.

Harris, D., & Williams, J. (2012). The association of classroom interactions, year group and social class. *British Educational Research Journal, 38*(3), 373-397.

Heritage, J., & Atkinson, J. M. (1984). Introduction. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: studies in conversation analysis* (pp. 1-16). Cambridge: Cambridge University Press.

Jadallah, M., Anderson, R. C., Nguyen-Jahiel, K., Miller, B. W., Kim, I., Kuo, L., Dong, T., & Wu, X. (2011). Influence of a teacher's scaffolding moves during child-led small-group discussions, *American Educational Research Journal, 48*(1), pp. 194-230.

Lefstein, A., & Snell, J. (2011a). Classroom discourse: the Promise and Complexity of Dialogic Practice. In S. Ellis, E. McCartney & J. Bourne (Eds.), *Insight and Impact: Applied Linguistics and the Primary School*. Cambridge: Cambridge University Press.

Lefstein, A., & Snell, J. (2011b). Professional vision and the politics of teacher learning. *Teaching and Teacher Education, 27*(3), 505-514. doi: 10.1016/j.tate.2010.10.004

Lefstein, A., & Snell, J. (2011c). Promises and Problems of Teaching With Popular Culture: A Linguistic Ethnographic Analysis of Discourse Genre Mixing in a Literacy Lesson. *Reading Research Quarterly, 46*(1), 40-69.

Lefstein, A., & Snell, J. (2013). Beyond a unitary conception of pedagogic pace: quantitative measurement and ethnographic experience. *British Educational Research Journal, 39*(1), 73-106.

Lefstein, A., & Snell, J. (2014). *Better Than Best Practice: Developing Teaching and Learning Through Dialogue*. Abingdon: Routledge.

Levinson, S. C. (1979). Activity types and language. *Linguistics, 17*(5/6), 356-399.

Matusov, E. (2007). In Search of the Appropriate Unit of Analysis for Sociocultural Research. *Culture & Psychology, 13*(3), 307-333.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, Mass: Harvard University Press.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359-377.

Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics, 1*(2), 137-168.

Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology, 80*(1), 1-14.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. London: Routledge.

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*(1), 95-111.

Molinari, L., Mameli, C., & Gnisci, A. (2012). A sequential analysis of classroom discourse in Italian primary schools: The many faces of the IRF pattern. *British Journal of Educational Psychology, 83*(3), 414-430.

Myhill, D., & Dunkin, F. (2005). Questioning Learning? *Language and Education, 19*(5), 415-427.

Nassaji, H., & Wells, G. (2000). What's the use of 'triadic dialogue'?: An investigation of teacher-student interaction. *Applied Linguistics, 21*(3), 376-406.

Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.

Schegloff, E. A. (2001). Discourse as an Interactional Achievement III: The Omnirelevance of Action. In D. Schiffren, D. Tannen & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 229-249). Malden, MA: Blackwell.

Selting, M. (2000). The Construction of Units in Conversational Talk. *Language in Society, 29*(4), 477-517.

Sinclair, J.M., & Coulthard, M. (1976). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press: London.

Smith, F., Hardman, F., Wall, K., & Mroz, M. (2004). Interactive whole class teaching in the National Literacy and Numeracy Strategies. *British Educational Research Journal, 30*(3), 395-411.

Snell, J. (2011). Interrogating video data: systematic quantitative analysis versus micro-ethnographic analysis. *International Journal of Social Research Methodology, 14*(3), 253-258.

Snell, J. & A. Lefstein (2011) Computer-assisted systematic observation of classroom discourse & interaction: Technical report on the systematic discourse analysis component of the Towards Dialogue study. *Working papers in Urban Language & Literacies*, #77. London: King's College London. <http://www.kcl.ac.uk/depsta/education/wpull.html>.

Varenne, H., & McDermott, R. (1998). *Successful failure: The school America builds*. Boulder, CO: Westview Press.

Vygotsky, L.S. (1987). *The collected works of L.S. Vygotsky: Vol. 1: Thinking and speech* (R.W. Rieber & A.S. Carton, Eds.). New York: Plenum.

Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., et al. (2009). 'Explain to your partner': teachers' instructional practices and students' dialogue in small groups. *Cambridge Journal of Education, 39*(1), 49-70.

Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge, U.K.: Cambridge University Press.

Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *Journal of the Learning Sciences, 15*(3), 379-428.

Wertsch, J.V. (1998). *Mind as action*. Oxford: Oxford University Press.

Yoder, P.J., & Symons, F.J. (2010). *Observational measurement of behaviour*. London: Springer.

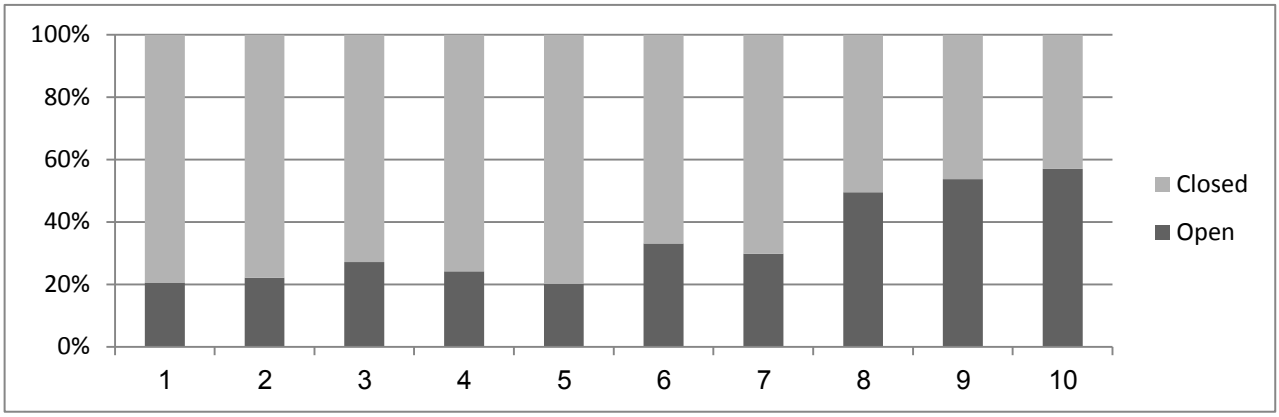


Figure 1. Change over time: open vs. closed questions

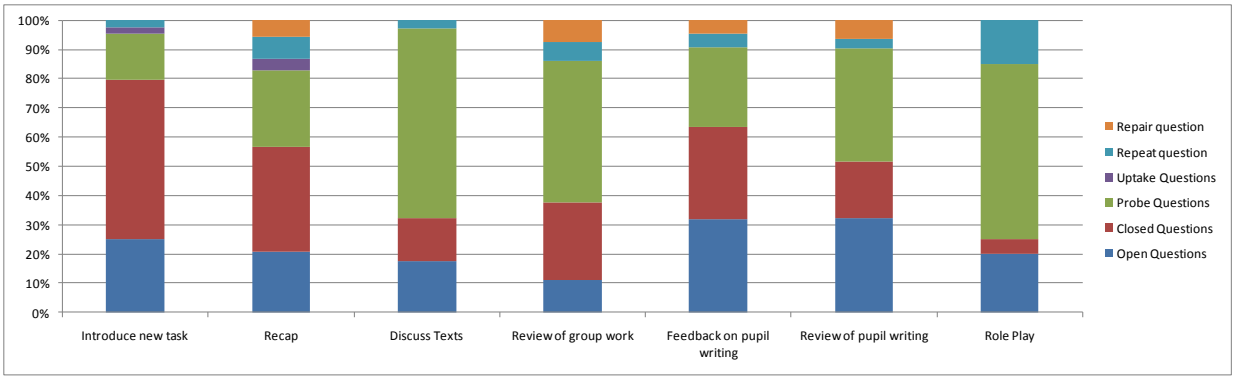
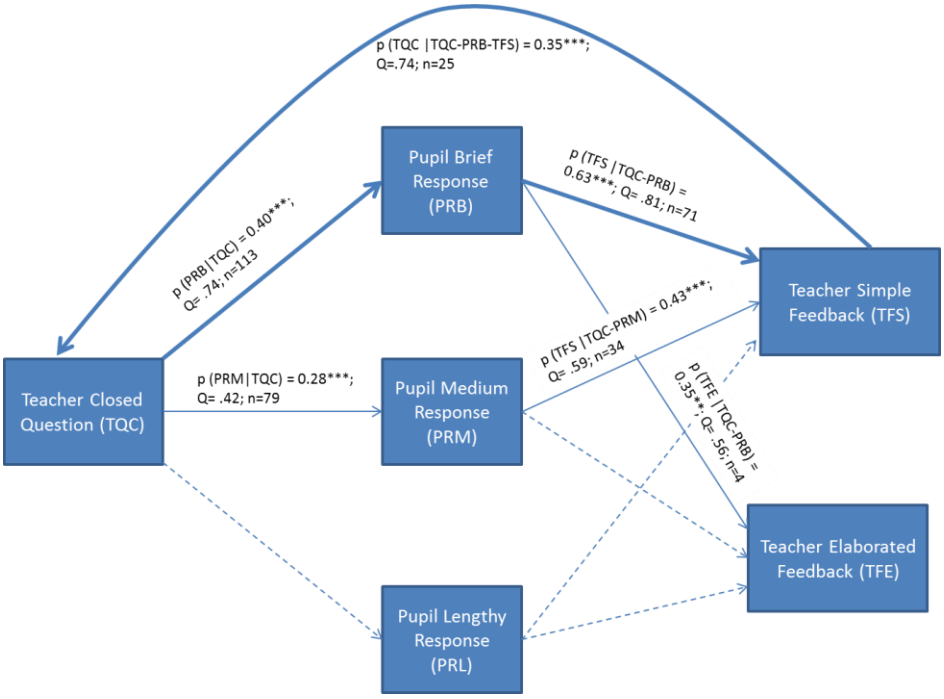
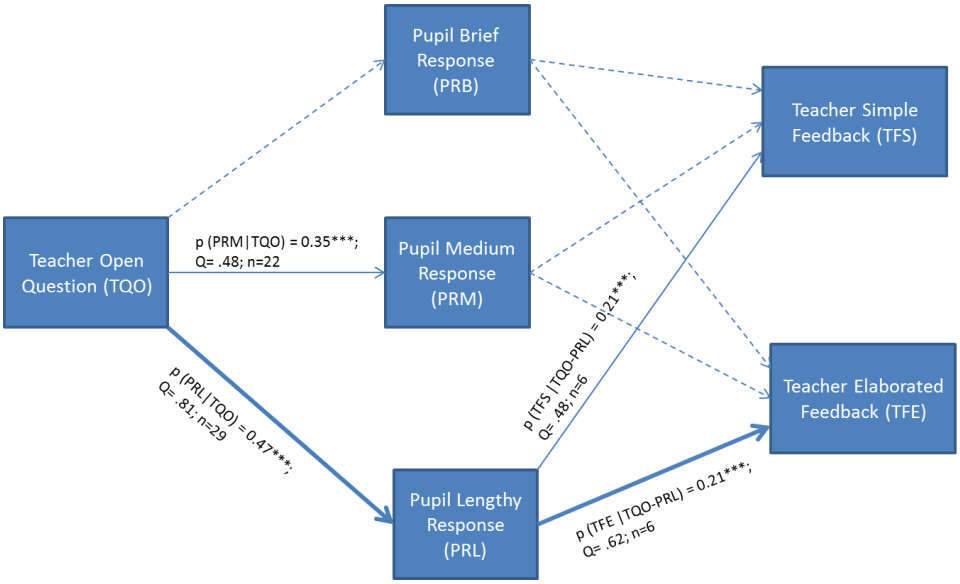


Figure 2. Discourse variation as a function of pedagogic activity in Ms Leigh's lessons



$^{**}p < 0.01$; $^{***}p < 0.001$

Figure 3. Sequential dependencies in chains initiated by TQC in Ms James' classroom



$^{**}p < 0.01$; $^{***}p < 0.001$

Figure 4. Sequential dependencies in chains initiated by TQO in Ms Leigh's classroom

	Ms Leigh		Ms James		Mr Robbins		School Average		National Sample
Lessons sampled	10		10		10		30		35
Total whole-class teaching sampled (mins)	224		262		244		729		
Mean lesson duration (mins)	49		50		46		48		53
Mean whole-class teaching duration (mins)	22		26		24		24		32
Mean percentage whole-class teaching	45%		52%		52%		50%		60%
RATE PER HOUR									
Teacher									
Explain	35		61		33		44		52
Direct	49		32		46		42		43
Open Questions	*20%*	16	*14%*	29	*18%*	19	*16%*	22	*13%*	15
Closed Questions	*30%*	24	*35%*	71	*36%*	37	*34%*	46	*50%*	58
Probe Questions	*38%*	31	*17%*	34	*26%*	27	*23%*	31	*17%*	19
Uptake Questions	*1%*	1	*4%*	9	*3%*	3	*3%*	5	*7%*	8
Repeat or repair question	*11%*	9	*30%*	62	*16%*	17	*23%*	30	*13%*	15
Total Questions	81		205		103		134		115
Elaborated Feedback	*42%*	26	*6%*	8	*6%*	5	*13%*	13		
Simple Feedback	*58%*	36	*94%*	116	*94%*	90	*87%*	83		
Total Feedback	**62**		**124**		**95**		**95**		**65**
PUPIL									
Spontaneous Contribution	21		19		49		30		10
Read aloud/pupil presentation	12		11		14		13		11
Response to Teacher	145		200		158		169		120
Choral response	19		52		15		24		9
Average duration pupil response to teacher (seconds)	6		3		6		5		5
Percentage contribution (duration)	32%		23%		41%		32%		25%

Table 1. Cross-classroom Comparison of Frequencies and Rates of Discourse Moves

Note: This is an abridged version of Appendix D in Snell & Lefstein (2011).

Code	Abbreviation	Count	Relative frequency
Teacher simple feedback	TFS	928	0.13
Teacher closed question	TQC	527	0.07
Teacher open question	TQO	248	0.03
Teacher elaborated feedback	TFE	147	0.02
Pupil moderate response (1.42-5.608 secs)	PRM	1084	0.15
Pupil brief response (0-1.419 secs)	PRB	549	0.08
Pupil long response (5.7+ secs)	PRL	547	0.08
Other	&	3198	0.44
Total		7228	1.00

Table 2. Frequencies of Discourse Moves across the Corpus

		Independence			Non-Stationarity between activities			Non-Stationarity over time			Heterogeneity		
Initi-ation	Class	LRχ^2	*df*	*p*	LRχ^2	*df*	*p*	LRχ^2	*df*	*p*	LRχ^2	*df*	*p*
TQC	Ms James	253.28	3	0.001	24.79	48	0.998	32.85	24	0.107	264.83	24	0.0001
	Ms Leigh	76.08	3	0.001	25.77	24	0.365	24.86	24	0.413			
	Mr Robbins	108.34	3	0.001	91.90	60	0.005	44.85	24	0.006			
TQO	Ms James	74.50	3	0.001	23.77	24	0.475	67.99	24	0.001	292.16	24	0.0001
	Ms Leigh	97.28	3	0.001	40.18	18	0.002	46.66	12	0.366			
	Mr Robbins	53.54	3	0.001	75.76	30	0.001	11.05	12	0.524			

Table 3. Likelihood Ratio Chi Square tests for Heterogeneity, Non-Stationarity and Independence

Chain	Ms Leigh		Ms James		Mr Robbins	
	Count	***z***	**Count**	***z***	**Count**	***z***
TQC-PRB-TFS (recitation)	5	0.46	71	6.85***	8	1.71
TQO-PRL-TFE (dialogic)	6	2.75**	0	-0.30	0	0.00

$^{**}p < 0.01$; $^{***}p < 0.001$

Table 4. Frequencies of Select Three-event Chains in Each of the Three Classrooms