

Coding classroom dialogue: Methodological considerations for researchers

Sara Hennessey, Christine Howe, Neil Mercer & Maria Vrikki¹

Faculty of Education, University of Cambridge, 184 Hills Road, Cambridge CB2 8PQ, UK

sch30@cam.ac.uk, cjh82@cam.ac.uk, nmm31@cam.ac.uk, mvrikk01@ucy.ac.cy

Learning, Culture and Social Interaction, 2020 (Authors' accepted version)

Highlights

- Systematic analysis or coding of classroom interaction is enormously complex and demanding
- Methodological challenges include scope, grain size, reliability and validity
- Adaptation of coding schemes for new research purposes is discussed and illustrated through a worked example

Abstract

Systematic analysis or coding of classroom dialogue is useful for assessing the role of high-quality interaction in supporting learning. However, although coding is an immensely complex and cognitively demanding activity that has taxed researchers over decades, the methodological challenges are often not discussed or problematised in empirical reports. Accordingly, this paper aims to help researchers make sense of the challenges, strengths and practical applications of using systematic coding schemes for analysing classroom dialogue. It presents an in-depth analysis of the pros and cons of contrasting approaches and the key methodological considerations, including scope, grain size, reliability and validity. It goes on to provide a worked example, illustrating how one team tackled the challenges in adapting for a new research objective an earlier coding scheme developed for use across diverse contexts. Two original, theory-informed analytic tools created to study the relationship between dialogic teaching and student learning in English primary schools are shared and made available for others' use or adaptation. The paper offers practical guidance for developing or adapting coding schemes for different research purposes. It highlights the need for further precision and critical attention to the ways in which scholars are investigating dialogic practices intended to support learning.

Keywords: dialogue, coding, analytic scheme

¹ Present address: University of Nicosia, 46 Makedonitissas Avenue, CY-2417, P.O. Box 24005, CY-1700, Nicosia, Cyprus

Coding classroom dialogue: Methodological considerations for researchers

1. Introduction

1.1 Theoretical underpinnings

There is a growing body of work concerning the features of classroom talk that are productive for learning (Resnick, Asterhan, & Clarke, 2015), namely, “dialogic” forms of interaction. These forms include posing open questions, participating in shared reasoning and thinking, and engaging critically but constructively with other perspectives (e.g. Mercer & Littleton, 2007). Closely related work identifies principles of “dialogic teaching” and “dialogic pedagogy”, namely the strategic use of different types of talk to achieve certain pedagogical goals, in ways that privilege dialogue (Alexander, 2008; Kim & Wilkinson, 2019; Skidmore & Murakami, 2016). Despite considerable consensus within the literature, a diversity of (overlapping) perspectives and emphases remains (Howe & Abedin, 2013; Schwarz & Baker, 2017), and researchers wishing to analyse classroom dialogue must begin by identifying the theoretical constructs that are pertinent for their work.

Research in this field often draws on a sociocultural perspective, that is one that emphasises learning through social and communicative processes. Children learn via problem solving, thinking and inquiry that is creative, open-ended and crucially conducted with others, and the quality of their interactions with other students and teachers, especially spoken interactions, is therefore paramount (Howe & Mercer, 2007; Mercer & Littleton, 2007; Vygotsky, 1978). Theories of dialogism indicate that speakers *position* themselves in relation to other participants, recognising diversity of voices, values, beliefs and perspectives (Bakhtin, 1981), and meaning emerges from tension between perspectives in that “dialogic space”; thus, it develops through and across *difference* (Wegerif, 2007).

Many researchers in the field focus on the collective, reciprocal, supportive, cumulative and purposeful nature of dialogue as described by Alexander (2008; see also Mortimer & Scott, 2003; Wells, 1999). This perspective includes construction of meaning through pursuing common goals and chained lines of collaborative inquiry in which answers give rise to new questions (Bakhtin, 1981; Kumpulainen & Lipponen, 2010; Wells, 1999). Researchers stress the importance of *connecting* to past/future events or wider contexts beyond the immediate interaction (Alexander, 2008) including reference to prior contributions or learning (Howe, Tolmie, Duchak-Tanner, & Rattray, 2000). They draw too on notions of guided participation (Rogoff, 1990) and scaffolding (Rojas-Drummond, Toerreblanca, Pedraza, Vélez, & Guzmán, 2013; Van de Pol, Volman, & Beishuizen, 2011; Wood, Bruner, & Ross, 1976) in explaining how dialogic interactions *contingently* adjust support and transfer responsibility to learners.

Learners are active rather than passive participants in the process of joint knowledge building through exploring, transforming, comparing, coordinating and analysing different ideas (Elbers, 1996; Mercer, 2000; Rogoff, 1990). They make influential responses and *extended contributions*, commonly through *elaborating, clarifying and building on* previous contributions made by themselves and others (e.g. Boyd & Markarian, 2015; Michaels, O’Connor, & Resnick, 2008; Nystrand et al., 1997; Wells & Arauz, 2006). Their dialogue also has a *metacognitive* dimension whereby teachers and students reflect on the quality and productiveness of their talk in achieving learning objectives (Fisher, 2007; Iiskala, Vauras, Lehtinen, & Salonen, 2011; Mercer & Dawes, 2008; van der Veen, de Mey, van Kruistum, & van Oers, 2017).

Importantly, the emphasis is not just on the open expression of ideas; hearing all voices is alone insufficient since ideas are not all equal. Exploring difference involves *critically evaluating* and challenging others' ideas and theories and the quality of the information and evidence/counter-evidence they propose. Thus, participants ideally go beyond simple statements of agreement/disagreement to engage in reasoned discussion and *coordination* of competing viewpoints and ideally, reasoned resolution (although this may be deferred: Howe, McWilliam, & Cross, 2005). *Reasoning* plays a major role in the relevant literature, for example in "exploratory talk", participants make their reasoning explicit to others and constructively question other perspectives (Mercer & Dawes, 2008). Similarly, Michaels et al. (2008) assert that speakers should be *accountable* not merely to the learning community, but also to knowledge and the standards of reason. In "Accountable Talk[®]", participants prioritise developing ideas and issues over presenting and defending their own positions. For Chinn, Anderson, and Waggoner (2001), pedagogic strategies supporting collaborative reasoning include scaffolding students to use text evidence to support their arguments and to challenge others, and praising those students who do challenge.

The *quality* or rigour of reasoning is not always finely distinguished, although there are exceptions. Webb et al. (2014) indicate that explicit, detailed and content-specific explanations are associated with learning gains. Likewise, work by Wilkinson and colleagues (2017) highlighting inquiry dialogue as an effective means of promoting students' argument literacy identifies an "evaluativist" epistemology which recognises the importance of students testing their ideas against others'. In their highly informative treatise on the history, theory and practice of dialogue and argumentation in education, Schwarz and Baker (2017) report that "deliberative argumentation" (Asterhan & Schwarz, 2016) – a kind of dialogue that integrates rigorous, structured reasoning and respectful reference to others – is the most productive form for learning. It models scientific talk but combines this with collaborative, interpersonal regulation and a democratic stance characterised by willingness to consider all – potentially conflicting – viewpoints and to change own position. The latter, inter-subjective dimension of learning draws on Bakhtin's (1981) dialogism; it is simultaneously accepting and critical. Evaluation of argumentation thus additionally encompasses the quality and equity of student participation (Schwarz & Baker, 2017).

1.2 Methodological approaches

Close study of classroom talk and its social and cognitive functions spans about 50 years, i.e. from the time that documentation became easily accessible through audio recording technology (Mercer & Dawes, 2014). An early boost to the field came from Sinclair and Coulthard's (1975) seminal identification of triadic Initiation-Response-Feedback structures (e.g. 'What is the capital of France?'), Response (e.g. 'Calais'), Evaluation/Feedback (e.g. 'No, it's Paris'), for it stimulated countless studies into the contexts where the structures are more or less probable. At the same time, the growing interest in productive dialogue encouraged researchers to address such issues as whether impressions of quality are grounded in actual instances, what forms of teacher questioning promote extended student responses, whether teacher development workshops support dialogic pedagogy and/or a wider repertoire of "talk moves"², how patterns of dialogue vary across school subjects, and whether increasing the quality of classroom talk promotes student learning.

² Talk moves are identified by their function within an interaction, signifying and supporting both social and cognitive processes. For example, a move can involve both reasoning and listening and responding to others (Greeno, 2016; Resnick, Michaels, & O'Connor, 2010).

Key methodological approaches are collated and illustrated by Kershner, Hennessy, Wegerif, and Ahmed (2020), and include linguistic ethnography, where talk is viewed as negotiating identity (e.g. Lefstein, 2008; Maybin, 2006), sociolinguistics, which stresses the importance of examining the surrounding social context (e.g. Chinn et al., 2001; Gee & Green, 1998) and ethnomethodology, where talk is primarily explored through conversation analysis with a focus on sequences of moves, turn taking and positioning of participants (Garfinkel, 1974; Sacks, Schegloff & Jefferson, 1974). Sociocultural methodologies arising over the last couple of decades typically build on these approaches when studying the discourse activities through which knowledge is socially constructed. They take account of the cultural and historical context and practices that shape the meaning of teachers' and students' contributions (e.g. Gee & Green, 1998). Sociocultural discourse analysis (Mercer, 2004), elaborated further below, introduces methods of quantitative analysis alongside the traditionally qualitative focus in order to capitalise on the relative strengths of these approaches.

No matter which approach is followed, there will come a point where aspects of the interaction are categorised or "coded". Coding may mean little more than identifying and interpreting some aspect of practice that is present in (only) some exchanges, but as with all other aspects of human social behaviour, it is an inescapable element of research into educational dialogue. Coding may reduce the unknowable reality to what can be easily measured (Wegerif, 2020; Wegerif & Mercer, 1997), but just as ordinary human interaction is inconceivable without such reduction, so it is part-and-parcel of any research process. Thus, decisions about the approach to coding are critical, yet these are seldom discussed in the literature. Published reports rarely report on the trials and tribulations involved; indeed, readers could be forgiven for assuming that the processes of development and application had been quite straightforward. Recognising this, the main aim of this paper is to illuminate and problematise the central methodological issues arising in developing or adapting coding schemes that relate to educational dialogue, for the benefit of researchers grappling with the inherent challenges. Our focus is upon spoken dialogue, although it is recognised that there are other modes of educational dialogue.

We start by discussing such generic issues as the utility, appropriateness and adaptation of coding schemes for different research purposes, the scope and granularity of units of analysis, and the achievement of reliability and validity. We then introduce our own adaptation of a coding scheme (*Scheme for Educational Dialogue Analysis* or SEDA: Hennessy et al., 2016) as a worked example of decision making. We describe how SEDA was subsequently reformulated and condensed into a new version (the *Cambridge Dialogue Analysis Scheme* or CDAS: Vrikki, Wheatley, Howe, Hennessy, & Mercer, 2019) and used by some of its originators for a new research purpose: a large-scale investigation of the relationship between teacher-student dialogue and student learning and attitudinal outcomes (Howe, Hennessy, Mercer, Vrikki, & Wheatley, 2019). Here we share our approach with others who may be interested in developing, adapting or critiquing coding approaches, with the aim not of offering any kind of recipe but of sharing reflections on the strengths and boundaries of approaches to coding in general and on the development of our own tools.

2. Developing or adapting a scheme for coding: Methodological considerations

In this section we explore and illustrate some key methodological issues arising with the initial development of any systematic coding scheme or with the subsequent adaptation for new

purposes.³ Tschan, Zimmerman, and Semmer (2018) warn against developing a new scheme unless necessary, because of the lengthy process entailed and the difficulty of comparing research outcomes with previous studies. At the same time, they warn against taking a scheme off the shelf unless it is perfect for the purpose. This leaves researchers with a real challenge. Adaptation of any coding scheme means that categories will almost certainly need to be selected and quite possibly reformulated to address the specific research emphasis. For example, the well-established typology of productive questions characterised by Nystrand, Gamoran, Zeiser, and Long (2003) using the Classroom Language Assessment System, CLASS 2.0 (Nystrand, 1988) identified three key discourse moves by teachers and coded interactions surrounding these: 1) authentic questions – open, with no predetermined answers; 2) uptake – previous answers are incorporated into subsequent questions; and 3) high-level evaluation – teachers elaborate students' replies or ask follow-up questions. Questions from students proved especially effective for learning in secondary English and social studies classes. This scheme has been drawn upon and radically extended by subsequent researchers in the Quality Talk tradition (Wilkinson, Soter, & Murphy 2010), where additional categories have been included in order to expand the analysis for new research objectives. Soter, Wilkinson, Murphy, Rudge, Reninger, and Edwards (2008) sought to develop a comprehensive model of quality discussion that promotes high-level comprehension of texts and included broad-grained categories such as elaborated explanations and exploratory talk, along with specific reasoning words. Their adaptation of Quality Talk encompassed changes to *scope* – what the scheme actually covered – and mixed the levels of granularity; some categories applied to individual questions and others to sequences.

2.1 Scope

The first decision to be made relates to scope; some schemes are designed for analysing all forms of discourse, not just dialogic forms (e.g. Boyd & Markarian, 2015; Wells & Arauz, 2006). These are maximally comprehensive but whether this is appropriate depends on the research purpose. Researchers coding only dialogic forms need to be sensitive to the implications of leaving gaps. Moreover, every scheme must have some boundaries. For instance, only some, not most, researchers in this field address the affective dimension of interaction (Pianta & La Paro, 2003) or infer meaning from “contextualisation clues” (Gee & Green 1998, p.122) such as gesture, gaze and tone. In our own studies gesture and gaze are drawn on to clarify references that may otherwise be ambiguous, and tone within recorded speech is likewise taken into account when the function of utterances is unclear, but these are not categorised as such.

Some schemes relate to teacher discourse moves only, for instance Accountable Talk® (Michaels et al., 2008) and the Analyzing Teaching Moves [ATM] Guide (Correnti et al., 2015). Other schemes address student moves only, as in the 12 forms of extended contribution identified by Hardman (2019) and the Functional Analysis of Children's Classroom Talk [FACCT] used by Kumpulainen and Wray (2002). Further schemes code both student and teacher moves, but approach these differently (Boyd & Markarian, 2015; Osborne et al., 2015). Sedova, Sedlacek, and Svaricek (2016) coded four indicators to assess change after their professional development intervention and these were mixed as follows: student talk with reasoning, teachers' open questions of high cognitive demand, teacher uptake, and open discussion. Regardless of the choices made, it is clearly essential to identify teacher and student moves separately since certain types of contribution (e.g. those performing a guiding or regulatory function, or synthesising ideas)

³ A substantial methodological handbook on group interaction analysis edited by Brauner, Boos, and Kolbe (2018) offers a wealth of further technical detail.

probably carry more weight when made by students than when made by teachers. Some researchers may wish to go even further and follow contributions of individuals over time.

Several researchers focus on the third turn in triadic dialogue or the “follow-up” move, widely considered to be significant in shaping (or obstructing) the course of dialogue (Park, Michaels, Affolter, & O’Connor, 2017). This move can of course stimulate another exchange, and so on (see Rojas-Drummond et al., 2013, for discussion of spiral, dialogic cycles of Initiate-Response-Feedback). The ATM Guide for coding whole class discussion (Correnti et al., 2015) likewise distinguishes teachers’ Rejoinder moves (e.g. Repeat, Uptake, and Connection between ideas or to prior learning/interaction) but also includes Initiating moves (e.g. Launch – an open-ended question; Literal – a factual question; Provides Information). In peer dialogue, distinguishing initiating versus responsive moves by learners can yield important information about the degree of symmetry/asymmetry (Maine, Rojas Drummond, Hofmann, and Barrera, 2020).

2.2 Granularity

The next decision concerns the size of chunks to be analysed, i.e. their granularity. In *The Ethnography of Communication*, Hymes (1972) proposes a set of hierarchical and nested levels of systematic analysis to investigate conversational sequences in diverse sociocultural contexts (see also Saville-Troike, 2003). “Communicative acts” lie at a micro level, defined by the minimum number of utterances or actions needed to reflect their interactional function and obtaining their status from the social context as well as their grammatical form and intonation. Communicative acts are embedded within “communicative events” at a meso level, defined by changes in participants, purpose, task, orientation and/or general topic. These in turn are part of broader “communicative situations” at a macro level (e.g. lessons). The levels are not mutually exclusive in that sequences can be analysed simultaneously at several levels, and indeed this was the approach that Hymes himself followed. Nevertheless, much research in the field opts for one level only, with the micro-level by far the most popular.

2.2.1 Micro-level coding in the analysis of classroom dialogue

The popularity of micro-level coding stems partly from the nature of the key concepts: open questions, extended contributions, reasoning with evidence and so on are all micro-level *constructs*, so to the extent that these are the research focus it is only natural to employ micro-level codes. Additionally, many research questions require a quantitative approach, and micro-level coding provides a straightforward route to quantification since frequencies can be counted. Advantages of frequency counting include being able to: 1) reduce and process large quantities of data to highlight key markers of dialogue; 2) search the dataset efficiently and see in close detail how specific acts manifest themselves and correlate with each other; 3) detect turn-taking and other patterns; 4) measure change in practice or learning or student participation over time; 5) characterise the variety of forms of dialogue within a specific educational setting; 6) make comparisons (e.g. across groups, classrooms, schools). For instance, the professional development intervention by Alexander et al. (2017) measured changes in the frequencies of brief versus extended contributions, recitation versus discussion/dialogue, closed versus open teacher questions. Researchers also look at ratios of dialogic compared to non-dialogic forms of interaction, as undertaken in our own study (Howe et al., 2019). In addition, statistical techniques can be used to determine the association of certain types of dialogue or frequency of contribution

with learner characteristics (e.g. socioeconomic status or linguistic capability) or achievement and attitudinal measures (ibid.).

Counts have also been made of the relative frequencies of particular words or patterns of language use (Mercer, 2004). For instance, one technique that Mercer, Dawes, Wegerif, and Sams (2004) used to assess primary children's reasoning and scientific understanding involved counting the frequencies of key words ("because", "if", "I think", "would" and "could"), long utterances, and links and labels in pre- and post-intervention concept maps. All of these measures predicted positive learning gains, so the study offered statistical evidence of changes in both group and individual reasoning. Specialised concordance software such as AntConc or Wordsmith assisted the process by enabling text files to be scanned rapidly for target instances. Thus, the relative incidence, repetition and form of use could be compared across contexts, and which words tended to co-occur (collocations) was revealed. Collocations and repetitions can reveal some of the more subtle, local meanings that words have gathered in use (ibid.). Hypotheses can be generated through pilot work and then tested systematically across a larger corpus of text, highlighting how teachers and students use particular terms in relation to their joint activity. Our own team has recently explored this technique, in conjunction with bottom-up qualitative analysis (Hennessy, Mercer, Calcagni, Leung, & Lim, 2019).

A further advantage of working at the micro-level is that it allows researchers to conduct fine-grained, systematic analyses, which will not only keep as close as possible to what participants actually say and do but arguably also allows maximum flexibility (Brauner, 2018). Any category definition can be broadened or broken down further as desired. For instance, analyses conducted by Hardman (2019) after a large-scale pre- to post- dialogic teaching intervention (Alexander, Hardman, Hardman, Rajab, & Longmore, 2017) revealed more extensive repertoires of teacher and student talk moves; one key move – *extended student contributions* – was further broken down into 12 smaller component talk acts such as *rephrase*, *recount* and *evaluate*. Wells (2001) categorises invitations for clarification in terms of requests for *repetition*, clearer *identification* of what the speaker intended to refer to, and *confirmation/denial* of the truth or validity of a previous utterance, whereas many researchers including ourselves simply include a category of *invitation for elaboration or clarification*. By contrast, while most researchers separate these invitations from those for reasoning, the PRACTISE argumentation rubric for teacher development developed by Osborne et al. (2015) combines *press for elaboration* with *press for explanation/evidence*. The rubric's purpose is, of course, informing teachers about ways in which they can stimulate argumentation by students so the broad-brush approach may work well.

Indeed, in larger coding schemes, related micro-level categories are usually grouped under headings or *clusters*, which not only preserve the broadening and breaking down but also permit independent usage in analyses. For instance, the sub-categories of Accountable Talk® are linked to teachers' goals (or challenges) in orchestrating whole-class talk that supports reasoning and robust learning (Park et al., 2017). The top-level categories in Wells' (2001) complex scheme are presented under headings such as Activity Orientation, Episode Development and Student Link. As well as acting as an organisational tool, clusters can help to maximise reliability (discussed in Section 2.3). Moreover, offering fewer, broader grained clusters is useful for time-poor practitioners. Researchers need to make their own decisions about whether and how to cluster, and these may be formed through deductive processes, i.e. derived from the literature.

2.2.2 Challenges of micro-level coding

Nevertheless, despite its advantages, micro-level coding is not without its challenges. For one thing, the concept of “communicative acts” and hence the micro-level is in itself multi-layered, embracing clauses, sentences, utterances, and turns. The turn level is a popular unit of analysis since changes of speaker (including changes caused by interruptions) can be fairly reliably distinguished by coders whereas the boundaries of an utterance may be more ambiguous. However, teachers sometimes address multiple students in a single turn and therefore decisions about segmenting within turns are often needed; this affects not only code frequencies but also turn length measurements. Moreover, a long turn includes multiple utterances and so, for more detailed coding, the utterance may be a more appropriate unit, as Park et al. (2017) argue; an utterance has a unique interactional function and a particular linguistic form. At the same time, focusing on utterances raises issues about whether to apply the same code repeatedly when substantially identical utterances recur within some turn, or whether to use a binary (present/absent) distinction. In principle, a quantifier can be added to binary coding, such as low, medium or high intensity, although this clearly complicates actual decision making. Where an utterance is interrupted by another speaker and continues in fragments, as is common in peer dialogue, researchers normally apply a code once only if frequency counting is envisaged.

A specific illustration of the turn versus utterance dilemma appears in our recent analysis of the implications of school subject for the relation between dialogicality and student attainment (Amodia-Bidakowska & Hennessy, 2018). We coded our large dataset of lesson recordings at the turn level, and calculated the number of turns per lesson that were assigned each code at least once. However, we then observed that average turn length in English lessons was significantly longer than in mathematics lessons. This did not matter for examining the association between dialogic interaction and attainment *within* subjects; while the range of frequencies was thereby smaller within English, the natural variation between teachers still emerged within each subject and the relationship to outcomes could be measured. It could be argued, however, that classrooms where participants take extended turns have lower frequencies of all codes and so the number of coded turns, regardless of length, is being privileged over the proportion of speech that is coded as dialogic. Since we were also interested in how the distributions of codes varied *across* subjects, English lessons could be unfairly treated if we simply compared raw frequencies across subjects for each code would likely appear less often in English, artificially deflating them relative to mathematics. So we corrected frequencies for varying numbers of turns/subject to elicit the *proportions* of lesson turns coded in each subject. Fortunately, our initial observation that more reasoning took place in mathematics lessons in fact held up and overall patterns of subject differences in the proportions of turns coded mirrored the results generated from raw frequencies. Similar corrections could be applied to any unit of analysis where the danger of over- or under-representation of codes arises, since unlike with our data, outcomes may differ.

Mercer (2010) highlights further challenges with micro-level coding, of which “the most serious are the problems of dealing with ambiguity of meanings, the temporal development of meanings, and the fact that utterances with the same surface form can have quite different functions” (p.4). The loss of the temporal dimension is especially acute when micro-level coding is coupled with frequency counting, since this removes all sense of how moves co-occur and are chained within exchanges (e.g. Lefstein, Snell, & Israeli, 2015; Wells, 2009; Wells & Arauz, 2006), or of how participants in dialogue *respond* to each other's contributions, which of course is central to any dialogic interaction (Bakhtin, 1981). Although many researchers do attend to the wider context of individual speech acts when assigning codes, they do not typically cast their net further back than immediate precedents or current frames. Yet while capturing dialogue over time is desirable, examining sequential contingencies is also very demanding. Moreover, context dependence leads

to variation in the quality of dialogue across different lessons as well as classes and means that analyses benefit from being repeated over time rather than taking a snapshot approach.

Mercer's further point about form-function relations is echoed by Boyd and Markarian (2015, p. 517) who write that "form follows function, not the other way around". For instance, an open-ended question does not necessarily stimulate an elaborate answer (Sedova et al., 2016). It is actually uncommon to characterise an initiating act (e.g. invite explanation) according to subsequent take-up or response (e.g. whether an explanation was provided) since it is the *dialogic intention or potential* that is usually deemed paramount (e.g. Wells, 1999). Intention has many facets, including the actual act of saying something, the function performed, the intended effect on the audience, and the actual effect (communication can misfire). In other words, there are many potential forms of interaction that can be used to accomplish a particular function, and similarly, multiple meanings or purposes can be communicated by identical words (Park et al., 2017). Analytic techniques need to attend to the range of possible permutations.

A special issue of the journal *Qualitative Inquiry* edited by St Pierre and Jackson (2014) argues that coding processes (and seemingly micro-level coding in particular) can be atheoretical or dogmatic in their conclusions. Such processes can assume a realist ontological position and may fail to recognise the personal influences of the researcher's own perspective and familiarity with the research context in which a scheme was derived, and the myriad of possible coding frameworks that could apply to the same dataset. "Diffractive reading of data", for example, aims to generate multiple theoretical insights and new kinds of knowledge through plugging data into theory into data as they constitute each other, thereby extending thought and theory in new directions and unpredictable patterns (Mazzei, 2014). The authors in this issue call for "post-coding" analysis that is open, critical and never-ending, unreplicable, unwieldy and subject to constant modification and directional changes rather than a linear process, and that is generative of new questions.

Sociocultural discourse analysis could be construed as a constructive example of post-coding analysis. Here categories are generated as outcomes rather than predetermined. The actual talk remains accessible throughout and is open to iterative scrutiny (Mercer, 2004). Deep analysis of participants' underlying intentions and responses to others moreover enables the researcher to go beyond the data, for instance to identify missed opportunities for extension or challenge. However, this is not the purely inductive approach it appears; Mercer and colleagues offer typologies of common forms of teacher-student and peer talk related to their own research purposes (the quality of student talk and how it is influenced by the teacher). These are not used for frequency counting but offer a useful frame of reference and heuristic device: the researcher has them in mind as s/he views the interaction yet remains open to new observations. The outcome – through team review of video recordings over time – typically includes narrative commentaries on specific episodes and comparisons between them (including lessons in the same classroom over time), with reference to prior knowledge of both the field and the context. This detailed qualitative analysis of talk within specific events can be combined with comparative analysis across a representative sample of cases, using a quantitative method. The criteria for choosing extracts may not always be clear, however.

2.2.3 Complementary methods

Many of the problems highlighted above stem from the use of micro-level coding in conjunction with frequency counting. However, micro-level coding can be employed with alternative quantitative techniques or without any quantification whatsoever. Nystrand et al. (2003) provide a

compelling example of the former in their use of event history analysis⁴ to study the antecedents and consequences of teachers' and students' questions and to investigate which question types precipitate dialogic sequences within lesson episodes. Detailed information on the number, types, and sequencing of questions posed by teachers and students was combined with assigning a 'dialogic value' to questions. These values were plotted as they unfolded over time, and the densities of emerging clusters of questions forming dialogic and monologic spells, and shifts between them, were examined. Event history analysis showed that dialogic discussion was significantly more probable when preceded by high proportions of student questions, uptake, and questions with high cognitive demands. The researchers point out that the patterning and sequencing of these elements, not their global averages, were crucial; analytic techniques allowing capture of these temporal sequences are thus very useful.⁵

Computer-assisted techniques, reviewed by Glüer (2018), are valuable not only for time sampling and contingency analyses but also when the aim is sequential analysis of chains of interaction moves, again addressing the temporal challenge. "Lag sequential analysis" calculates the probabilities of certain moves following other moves, especially questions; this requires a very large dataset (but a manageable number of categories) so will not be feasible for many researchers but offers insight into how closed questions are typically followed by minimal pupil responses and then simple teacher feedback whereas open questions stimulate more elaborated responses (Lefstein et al., 2015). Rich, dialogic chained sequences include teacher moves more characteristic of uptake than simple evaluation, resulting in a more exploratory stance towards the topic and more equitable participation (Molinari, Mameli, & Gnisci, 2013). A promising new technique, "epistemic network analysis" (Shaffer, 2017), integrates in-depth, qualitative analysis with statistical tools to model the connections made by each participant between an expressed idea and those in its recent temporal context.

Whether or not quantification is employed, it can be helpful to approach micro-level coding from several angles, perhaps approximating the "diffractive reading" discussed above. For instance, given the situatedness of speech acts and the ways in which the context can affect their interpretation by interlocutors (Edwards & Westgate, 1994), in-depth qualitative analysis of exchanges that takes account of the sociocultural context, other classrooms, and theory within the field, can supplement any initial coding. Indeed, many researchers concur with Gee and Green (1998, p.120) that systems for analysing classroom discourse must be integrated within an ethnographic perspective, seeking ways of capturing the role of wider educational and cultural practices in shaping the meaning of contributions to dialogue. This calls for thick, multimodal descriptions of discourse data (e.g. Lefstein et al., 2015; Wells, 2001); these may complement or replace micro-level coding. Rigour within such narrative analysis is assisted by backing up the interpretations with close reference to the data excerpts and seeking counterexamples; it may also involve consulting with participants themselves (Lefstein et al., 2015). Coding outcomes can also be related to factors such as activity type or subject culture; for instance, our team is currently exploring whether investigation/inquiry activities are productive for student questioning and reasoning (Amodia-Bidakowska & Hennessy, 2018).

⁴ This quantitative analysis technique permits addressing both fine-grained and broader questions, such as why might only certain classes shift from monologic to dialogic patterns of discourse? Exactly what teacher and student moves allow this shift? (Nystrand et al., 2003, p. 141). Longitudinal data were investigated using logistic regression, examining whether or not the events in question – dialogic spell, student question, or discussion – had occurred in a given classroom episode, and their timing.

⁵ Built-in software timestamp options for transcription and coding are very useful here.

Reference to wider educational and cultural practices implicitly reintroduces Hymes' (1972) three levels, and the micro-, meso- and macro- distinctions, with which this section began, for representing these practices must call in effect on meso- or macro-codes. The implication is that the ideal coding scheme will operate on several levels. However, the situation is complicated yet further through the fact that some of the features of productive dialogue pinpointed in Section 1.1 actually lie on one or other of the higher levels. Examples include Mercer's (2004) concept of exploratory talk and Alexander's (2008) five principles, both of which amount to codes at the meso-level. Thus, meso- or macro-level coding may not only be desirable to support micro-level coding; it may also be required to represent certain key features.

Up to a point, similar dilemmas apply when coding at the meso- or macro-level as were highlighted earlier with micro-level coding. For instance, paralleling the utterance versus turn uncertainties discussed above, meso-level concepts like exchange, episode, topic and sequence have proved controversial. The boundary of an episode is characterised by Sedova et al. (2016) as a change of activity, theme or communication approach, but by Linell (2001) as a bounded discourse event that is usually focused on the treatment of some problem, issue or topic (a matter of continued, sustained salience or importance) and maintains a single participant framework. Similarly, Wells (2001) defines an episode as a change of task or participant structure while Nystrand et al. (2003) regard it as a change of purpose or topic. Furthermore, several topics may be simultaneously open and intertwined within an episode according to Linell (2001). Wells defines a sequence as a nuclear exchange (two or three moves) and all exchanges that are bound to it through their meaning and function. Two or more successive exchanges (each made up of one or more moves) form a transaction according to Sinclair and Coulthard's (1975) original model of classroom talk. Lefstein et al. (2015) assert that interactional structures of three or more turns constitute the optimal unit of analysis. Shaffer (2017) found that 98% of recorded interactions referred to contributions within the preceding five lines, and in contrast with those who segment according to criteria such as shift in topic, segments discourse at the meso-level into "stanzas" – overlapping sequences of four to seven lines – for his epistemic network analyses.

At the same time, it is important not to over-stretch the parallels between micro-level codes and their meso- or macro-counterparts. For one thing, all forms of meso- and macro-level interaction allow us to make some sense of linked chains (Mortimer, 2005), taking into account the context that frames each chain (Myhill, 2006). For another, coding at the meso or macro levels may be best grounded in a continuum-based model rather than one that simply records the presence/absence of target forms. This is not purely a methodological issue but also informed by theory. For example, the view of many researchers in the field including ourselves is that efficacy of dialogic practices in terms of supporting learning is dependent on creating a *supportive classroom ethos* whereby teachers hold a dialogic intention and stance (Boyd & Markarian, 2015; Wells & Arauz, 2006) and students have increased agency and participation. A dialogic pedagogy entails investing time in developing norms and protocols for dialogue based on mutual respect and trust whereby learners feel comfortable in openly expressing tentative ideas and different viewpoints, engaging in trial and error and changing their minds in light of input from others (e.g. Hofmann & Ruthven, 2018; Kershner, 2020; Nystrand, Gamoran, Kachur, & Prendergast, 1997). Capturing the degree to which this is successful necessitates moving towards a more macro-level analysis and one that includes the affective and ethical characteristics of classroom communication, alongside the cognitive elements. Rating scales in particular can introduce a measure of quality that can distinguish optimal practice from more superficial student involvement. An example of one such scale is described in Section 4.3. A further example is a set of ten 3-point scales used to rate groupwork quality within each (sampled) episode of small group

work observed in our recent large-scale study (Howe et al., 2019) and adapted from a scale previously developed and tested by Howe et al. (2007). As well as frequency of dialogic moves, items covered cohesiveness, equitable participation and positive attitude towards collaboration; for example, “All pupils were involved in the group work interactions” and “Pupils tried to reach consensus or compromise”.⁶

2.3 Reliability and validity

Reliability testing is a common concern among developers of coding schemes who attempt to minimise inference levels and maximise the chances of two or more coders applying the same category to each unit of analysis. A number of indices are available for checking reliability, with Cohen’s Kappa most commonly used. The value of testing is questioned by some (e.g. Lefstein et al., 2015; Sedova et al., 2016), especially the implicit expectation that researchers unfamiliar with the context can code reliably using a scheme they have not developed. If coding is to be confined to one team and the scheme is not intended for general use, consensus coding may be used (e.g. Elizabeth, Ross Anderson, Snow, & Selman, 2012), although this is rare. Yet if the scheme, or indeed the dialogic features it is meant to address, is intended for general use, perhaps especially by practitioners, it is important that it can be communicated. While high reliability does not guarantee communicability, low reliability definitely precludes this.

In practice, levels of success vary with classroom dialogue and achieving high reliability levels is notoriously difficult and time consuming, typically taking up to 6 months or even more for a complex scheme. The finer grained and more complex a coding scheme, the less reliable it is likely to be simply because there is more room for discrepancy and error. Levels of inference and potential over-interpretation need to be monitored carefully, and a detailed coding manual should be developed – ideally with illustrative examples covering inclusion of ambiguous cases – and regularly consulted and updated. Reliability test results cannot be assumed to generalise to the rest of a dataset, except when there are very frequent codes and large test samples (Eagan et al., 2017). Researchers employ various strategies to maximise reliability, including tightening up and illustrating definitions, coding at cluster level, and using sequential, hierarchical structures of decision making. For example, Transactive Discussion (Berkowitz & Gibbs, 1983) begins with a distinction between orientation towards self or other and then breaks down those categories.

Unlike reliability, construct and content validity are very rarely addressed in this field. Researchers seem to presume knowledge of productive dialogue when formulating schemes to analyse it. The variation in definitions and markers used across studies (Howe & Mercer, 2016; Kim & Wilkinson, 2019), indicates that further work is needed to sharpen our understanding. One approach is theoretical triangulation, whereby the same transcript of discussion is analysed from different theoretical perspectives, yielding deeper, complementary insights from each (Clarke, 2001). Van Drie and Dekker (2013) related the perspectives of teacher, students and domain to each other by, respectively, analysing interactivity of the discourse, conceptual level raising and historical reasoning, and integrating the outcomes. Comparison of coding schemes has also been made on occasion. Hennessy (2020a) applied three different schemes to the same transcript to explore their affordances: SEDA (see Section 3), Accountable Talk® (Michaels & O’Connor, 2015) and Transactive Discussion (Berkowitz & Gibbs, 1983). There was strong overlap between the coding

⁶ Note that some of the scales were based on tallies made by observers using time sampling, i.e. on frequency counts of micro-level codes. This raises the question of whether some element of micro-level coding can ever be viably avoided. Ratings rather than frequencies could in fact be used for codes at all levels; they could apply to turns, episodes/events or lessons.

outcomes from the perspectives of the three schemes, plus some interesting, nuanced differences. Likewise, Mayer (2018) found that SEDA and Accountable Talk® moves could be roughly mapped onto her own Framing-Developing-Evaluating analysis of the history knowledge construction process, but a new category of “Asking for observations” was proposed. Further consideration is given to validity in Section 4.4, where the approach of expert validation is aired.

3. SEDA: The *Cam-UNAM Scheme for Educational Dialogue Analysis*

In the next two sections we describe how we adapted a complex micro-level coding scheme for a different research purpose, illustrating how the methodological considerations outlined above were addressed.

3.1. Introduction to SEDA

SEDA was first presented in an earlier article in this journal by Hennessy, Rojas-Drummond, and colleagues (2016) who developed and tested it iteratively over 3 years (a condensed version appears in Appendix A)⁷. It contains 33 categories characterised at the “communicative act” level (Hymes, 1972). Its scope mainly covers dialogic interaction, although the *Express Ideas* cluster addresses many otherwise uncoded aspects of classroom talk - its two, commonly observed components are not dialogic per se but are pivotal in sustaining dialogue. A fairly unusual feature is analysis of moves solely according to their function within the dialogue rather than to teacher or student roles as speakers, in order to recognise that these roles are potentially malleable (cf. Freire, 1996). Our own use of SEDA has involved segmentation at the communicative event (meso-) level as well as the micro-level and taking account of the wider sociocultural context in selecting codes and constructing an interpretive narrative.

SEDA’s 33 codes are clustered into eight groups: *Make reasoning explicit*, *Build on ideas*, *Invite elaboration or reasoning*, *Positioning and coordination*, *Connect*, *Reflect on dialogue or activity*, *Guide direction of dialogue or activity*, and the optional cluster *Express or invite ideas*. The theoretical constructs underlying dialogue and dialogic pedagogy summarised earlier are explicitly represented in these groupings (as the italic font in Section 1.1 conveys). The clusters underwent numerous iterations though. Ultimately, eight clusters proved sufficient to distinguish between key dialogic features to allow coding reliably at the cluster level; this served to increase practicability, additionally allowing for both use by practitioners and quantification (reliability of the 33 individual codes was of course much lower). Researchers can count frequencies more easily at the cluster level since codes within a cluster are hierarchical and thus mutually exclusive: only one is applied from each cluster to a single utterance. The final clusters are not clusters in any statistical sense but were formulated pragmatically to comprise theoretically related constructs and to maximise reliability. We explicitly acknowledged that many alternative permutations are possible; the current formulation served our purposes at the time, but it is to some extent arbitrary and is the third iteration we went through (as described by Hennessy et al., 2016).

In developing SEDA we recognised that the micro-level alone may be somewhat impoverished. A draft set of “global descriptors” was therefore devised to indicate environmental features that: 1) theorists consider to be supportive of dialogue (e.g. “Encourage free expression of opinions/ideas/beliefs”; “Explicitly situate oneself as co-learner” – including teachers); 2) describe

⁷ The full version of SEDA is freely available at <http://tinyurl.com/SEDAfull>

practices that foster dialogue skills (e.g. “Manage turns inclusively by shared routines” – including students); 3) describe longer exchanges rather than individual turns (e.g. “Extended questioning”). The rationale was that these should help to describe the character of dialogic practices – and to interpret sequences of moves – at a much coarser level of granularity than individual communicative acts. Some global descriptors proposed encompassed more than one function, e.g. “Exploratory talk” also used as a category by Soter et al. (2008), as mentioned above. Through grouping acts that occur in combination during episodes, these descriptors aimed to help recapture what is lost when dialogue is distilled down to micro-level descriptors. The global descriptors were never fully tested or published but a couple of them formed a useful basis for the development of whole-lesson rating scales (“dimensions”) as the second component of the CDAS package (described in 4.3). They could be developed further.

3.2 Design for adaptability

SEDA is an open resource with a Creative Commons license. It was explicitly designed for adaptability by other researchers for their own specific purposes and contexts. To maximise external validity when formulating the categories, our team attempted to be as comprehensive as possible in mapping and distilling the prominent theoretical perspectives in the field of educational dialogue, as outlined in Section 1.1. We argued that re-use and adaptation of such a scheme could contribute to building a common understanding and framework of terms for the central constructs within the field (Hennessy et al., 2016). We recognised of course that full coverage would be impossible owing to their diversity, and probably undesirable. However, this mapping aimed to provide a highly versatile tool that other researchers could use as a starting point for their own adaptations since they can be relatively confident that key constructs are not missing. Inclusion is only one step, of course, and the developer’s choices about emphasis and formulation inevitably determine the prominence and explicitness of any particular construct or approach.

While other researchers may find SEDA helpful in curtailing initial development, modification will almost certainly also be needed. For instance, while any SEDA category can in theory be applied to any participant, in practice, a researcher focusing on teacher moves or peer group dialogue may wish to shift the emphasis in some ways. Reclustering and even elimination of some clusters could well prove sensible in order to focus attention on constructs underlying the research question and to expand or condense the level of detail in relevant categories. The sheer number of categories may also need reduction for manageability. In developing our scheme, the deductive process of distilling key concepts from the literature was iteratively entwined with inductive refinement of draft categories through application to authentic examples from classroom data. Important for adaptability is that SEDA is widely applicable – and has already been tested – across age phases, subject areas, country settings and diverse interactional contexts including whole class, group and paired work.

SEDA has, for instance, been translated into Chinese, Hebrew and Spanish by other researchers and adapted for use in a variety of contexts. While one recent study used the original eight clusters to code 27 videotaped sessions in Chinese kindergartens (Huang, Yang, & Li, 2019), in most cases of re-use, there has been some adaptation to address diverse motivations. For example, Rojas-Drummond and colleagues have produced a streamlined version of SEDA (SL-SEDA) with five clusters comprising 18 communicative acts. They further modified this scheme for a study of peer dialogue whose aim was to distinguish symmetry and asymmetry of participation within small groups of (Mexican and English) primary school children reading and interpreting

different textual resources (Maine et al., 2020). The team prioritised utterances indicating high-level comprehension, structural asymmetries and “possibility thinking” in meaning-making. They collapsed some codes within and across clusters, yielding eight categories that are broader grained than SEDA codes (more akin to clusters, but without explicitly making fine distinctions within each category since these were not needed). A further layer of coding was needed to capture the nature of shifts made or attempted between episodes. In another study by Ludvigsen, Krumsvik, and Breivik (2020), three SEDA codes (alone) were used to operationalise exploratory talk when analysing peer discussions among undergraduate students.

Mayer (personal communication) tested the boundaries of SEDA in a preliminary analysis of one teacher’s methods for framing content questions during discussions of a poem. She identified two areas where the SEDA framework needed to be enhanced for full coverage of her data: the use of material evidence, for example a teacher asking students to reference specific lines in a poem, and verbatim repetition. The latter involved closely attending to – and probing the meanings of – students’ own language. It was regarded as giving their contributions status and encouraging classmates to think carefully about what a student had just said, i.e. a form of “wait time”. We welcome further adaptation and input to the scheme.

4. An integrated micro- and macro-level package

In this section we describe our own recent adaptation of SEDA for a new research purpose, beginning with the development of a shorter micro-level scheme and moving to a set of broader lesson-length rating scales.

4.1 Development of CDAS (*Cambridge Dialogue Analysis Scheme*)

Our recent ESRC-funded project set out to investigate the relationship between dialogic approaches and student outcomes on national standardised tests, reasoning capability and attitudes to school and learning (<http://tinyurl.com/ESRCdialogue>; Howe et al., 2019). We wanted to identify *which specific aspects* of talk, if any, were demonstrably productive, since published research tends to make assumptions about this rather than testing it, and rarely distinguishes between different aspects. Hence a micro-analytic method was appropriate and speaker turn was the unit of analysis that was easiest and most reliable to apply in coding at scale. Rather than reinvent a new wheel, we capitalised on the extensive development work underlying SEDA and strategically modified it to meet our particular aim of capturing classroom dialogue where the teacher is present (with whole class, small group or individual). We used the new scheme to analyse 144 video-recorded lessons taught by 72 Year 6 teachers (students aged 10-11) in 48 state-funded schools in England. The diverse sample included schools in areas of high deprivation. Turn-level coding of *dialogue moves*⁸ during lessons in English, mathematics and science was used to create indices of dialogicality. Regression analyses further examined the relationships with student outcomes. Our application of the scheme to this dataset and the relations with outcome are reported by Vrikki, Wheatley, Howe, Hennessy, and Mercer (2019) and Howe et al. (2019) respectively. Here we focus on the development and testing of the tools created for the new purpose, the main one being CDAS, the *Cambridge Dialogue Analysis Scheme*.

⁸ Note the use of the term “dialogue” instead of “talk” moves and “communicative” instead of “speech” act; in our view dialogue can take place in other modes than oral, and can indeed be multimodal and, Hennessy (2011) argues, even nonverbal, as in technology-mediated environments.

One of the main motivations for adapting SEDA for our study was the pragmatic need for a scheme that while remaining meaningful could also be reliably applied across our large dataset of over 9000 minutes of recorded lesson video. While the eight SEDA clusters offered a possible framework, the latest incarnation did not quite suit our purpose, owing to an issue with mutual exclusivity, so we started with the 33 codes. Almost all of them were deemed relevant, however they were radically condensed and reformulated. Some were merged, one (“Propose action or inquiry activity”) was dropped, some were added, and others were redefined. Some categories remained largely the same: *Elaboration (EL)*, *Reasoning (RE)*, *Elaboration invitations (ELI)*, *Reasoning invitations (REI)* except that the distinction between own/other’s ideas was removed. In an authentic dialogue it is very difficult – and we argue, undesirable – to determine individual ownership of ideas since new ideas are cumulatively and jointly constructed (cf. Merleau-Ponty, 1968). The ELI and REI categories aimed to capture questions that provoke thoughtful or reasoned answers respectively. Similarly, the EL and RE categories capture extended responses that involve detailed descriptions or justifications. However, the emphasis on reasoned argument and coordination in the literature – the notion of accountability to knowledge outlined in the Introduction – led us to tighten the scheme in that direction. The detailed changes made and their rationale, structured according to the key considerations outlined in Section 2, are detailed in Appendix B.

In our two schemes, as with most schemes, some codes distinguish invitations from statements/propositions; this is the case with reasoning and elaboration in both schemes⁹, expression/invitation of ideas in SEDA, and coordination in CDAS. However, a few schemes combine invitations and propositions in a single code because empirical testing (e.g. Howe et al., 2000) has indicated that they are very low frequency and/or highly inter-correlated. In formulating CDAS we followed suit for such codes; for instance, *Reference back* includes inviting reference back.

CDAS was developed over a 5-month period (including input from colleagues at an open coding workshop) and then tested for usability and reliability, with further adaptations made over another 5 months. The final version of the scheme (see Appendix C) comprised 12 codes, mostly dialogic moves (apart from Agreement when occurring alone and Other Invitations). Since almost all of the original SEDA codes were incorporated in either the new CDAS micro-level codes or its macro-level dimensions outlined in 4.3, coding remained complex despite the reduced number of categories. It typically took a day per one-hour lesson in total, and more like 10 hours when coding direct from video using Elan¹⁰ software (although this saved on transcription time and costs).

4.2 Application of codes

CDAS codes were applied at the turn level in a binary fashion; this strategy maximised reliability. Codes were usually associated with different sentences within the turn, however if there was more than one move of the same type in a turn, we coded it once only. This increased reliability as assigning codes to (finer-grained) utterances led to inconsistency over determining boundaries of utterances and over whether a code should be reapplied to repetitions or the same communicative act be deemed to continue. We broke down a turn comprising several sentences,

⁹ Interestingly, in creating a reduced number of dialogue variables from CDAS categories for our statistical investigation of correlation with learning outcomes, we later combined the invitational and statement forms of each construct because they were found to be highly correlated; thus invitations for elaboration unsurprisingly led to more elaboration, and so on.

¹⁰ Eudico Linguistic Annotator: <https://tla.mpi.nl/tools/tla-tools/elan/>

or even a single sentence, into smaller units or phrases, if necessary, each allocated a line in the coding spreadsheet, applying two or more codes in sequence within a turn (Hennessy et al., 2016). Because segmentation is not always straightforward or reliable, we recommend that researchers make careful decisions about procedures here, measuring reliability of segmentation itself or pre-determining it before other coders engage.

In both of our schemes, codes from different groups could apply to a turn, but CDAS does not contain clusters in the same sense as SEDA, where assignation of categories was mutually exclusive within clusters. In CDAS the codes are merely grouped along four conceptually distinct dimensions for ease of use (see Appendix C) but the groupings have no practical implications in terms of code selection or reliability testing. By contrast, SEDA is roughly hierarchical (codes listed higher up within each cluster are prioritised) while CDAS allows any combination of codes to be applied to a turn. CDAS presents some rules for prioritisation between competing codes, however, as this was needed for frequency counting (regarded as the optimal approach to obtaining the quantitative indices that our regression analyses required). For example, if a single-utterance turn included both EL and RE, then this was coded as RE, since reasoning was regarded as higher order. To maximise reliability, *Refer to wider context* trumped *Reference back*. It should be noted that any prioritisation rules will affect frequencies, downplaying the role of those moves that are then ignored, but in our study such moves were very rare and will not have had any notable effect.

4.3 Macro-level analysis: Lesson-level rating scales

The limitations of micro-level coding outlined above highlight the benefits of complementing with broader categories, i.e. coding at the meso- or macro-level. We opted for the macro-level with CDAS, specifically the whole-class level, partly because during the development of SEDA, issues had arisen with achieving reliability of meso-level segmentation. This was especially the case when topics or tasks were revisited during a dialogue. A further motivation was the need to characterise features of interaction and ethos within dialogic classrooms that are extremely difficult to capture below the macro level. Here, we describe how our team derived macro-level codes that applied across whole lessons and that, in order to reflect the continuum nature of codes at this level (see Section 2.2), were expressed as rating scales.

The macro-level rating scales served to describe relative levels of student involvement and teacher direction along five dimensions (see Appendix D): 1) announcement/negotiation of *aims and objectives*; 2) dialogic or authoritative *monitoring and guidance* during group work; 3) degree of *reflection on learning process* or trajectory by students/teacher; 4) degree of explicit *focus on target dialogic practices* (e.g. “ground rules”); 5) level of *student participation* in expressing ideas publicly at length/succinctly. These dimensions were created through various means. Although the 33 SEDA codes were intended for micro-level coding, a few could also characterise meso-level sequences. Our decision to transform these into whole-lesson scales elevated their status when frequency counts would otherwise have relegated them as infrequent, and it also enabled us to systematically capture key features of the classroom ethos as well as specific moves. One SEDA code in the Guide cluster, *Encourage student-student dialogue*, was simply transformed into the *Focus on dialogue* dimension, whereas others such as *Allow thinking time* were incorporated into descriptors; the whole *Reflect on dialogue or activity* cluster of three codes became the *Reflect on learning* dimension. A couple of the aforementioned draft SEDA global descriptors were also drawn on: *Manage turns inclusively by shared routines* and *Monitor participation in dialogue* were incorporated in the *Focus on dialogue* dimension.

Via these dimensions, we hoped to seek out and categorise a wide range of practices, such as those where students negotiate ground rules for participation, conduct genuine inquiry, critically evaluate others' ideas including debating and contesting authoritative perspectives, etc. We acknowledged that occurrence of the practices would depend heavily upon classroom history. For instance, where "ground rules" for talk had been carefully negotiated at the beginning of the school year and then successfully embedded and internalised by students, references to such rules would likely decline over time. Equally, we did not expect many students to define their own learning goals, redefining practices and values, as encompassed in Matusov's (2011) more radical notion of student "authorship" ("a participant's bid for a unique creative contribution fully or partially recognised [validated or contested] by a relevant community" (p.24). Nor did we imagine that many teachers would act as genuine co-learners with their classes, given their perceived need to address curriculum aims. Nevertheless, we explicitly allowed for these possibilities when the dimensions were expressed as rating scales.

For each dimension, ratings at one of three levels (0-2) were assigned once to each lesson – use of three points maximised reliability. The lowest level indicated that the dimension was not evident during the lesson, the middle level that it occurred but was predominantly teacher-led, and the highest level that there was student input. SEDA elements became part of the dimension descriptors, typically for the most dialogic level (rated "2"). Additional text was formulated to cover the potential degree of student involvement along each dimension, and exemplar practices were incorporated at each level. Actual recorded levels of dialogue according to these scales, and teacher-student differences are reported by Vrikki et al. (2019). Notably, further analyses revealed few strong associations between the micro-level dialogue categories alone and student outcomes, nor were the macro-level ratings alone predictive. However, the *combination* of two of our dialogue markers (*elaborating and building on ideas* and *challenge/disagreement with other perspectives*), plus high lesson ratings of *student participation* (active cognitive engagement by multiple students with others' ideas) did show a significant relationship in relation to two standardised tests (see Howe et al., 2019, for details). This has important implications, endorsing the value of multi-level coding.

4.4 Validation of the new tools

We have further investigated the validity and relevance of the CDAS package (and hence of SEDA on which it was heavily based) through soliciting independent evaluations from six selected experts whose views span a variety of analytic perspectives.¹¹ The experts rank ordered and intuitively commented on a sample of lesson transcripts from their own perspective in terms of how productive they considered the lessons for supporting children's learning. We identified the key criteria they used and explored the relationship between their written commentaries and the main categories within CDAS in order to check whether the categories represented key constructs that the research community would recognise and address the question of what, if anything, might be missed. The outcomes reinforced some of the strengths and limitations of analytic approaches to classroom dialogue. It was clear from the experts' comments that to a considerable degree, they used similar constructs to us; our main codes were mentioned to varying degrees (querying/challenging was mentioned less frequently) alongside some links between them. Those dialogic elements emerging as productive for learning featured heavily. One expert explicitly used student participation as a criterion in evaluating all of the lessons.

¹¹ Further details will be reported in a forthcoming publication by Hennessy, Mercer, and colleagues.

*Our research has shown that ... providing fully detailed, correct, and complete explanations, and **engaging with the details of others' ideas** (and teacher support of these kinds of interaction) are most predictive of outcomes.... We have applied this lens.*

Also related to our macro-level student participation scale was a broad theme emerging concerning levels of teacher direction versus student agency.¹²

One expert explicitly mentioned all three of our markers for learning through dialogue.

*[The teacher] is regularly seen **challenging their thinking** and allowing the children to **build on each other's ideas**. For example, we see her asking "why do you think that?".... do you agree and why, **inviting other children to add on** to what has just been said and asking for counter examples. Such moves allow the children to **elaborate on and develop their thinking**, and **play a greater role in the classroom talk as evidenced by their length of turns***

4.5 Inter-coder reliability testing of the new tools

In our study, systematic analysis of a very large dataset required a team of four coders to complete the work within the available timescale. All four coders were project team members and familiar with the context and aims of our study. Reliable coding across the team was utterly essential, with reliability testing carried out over a period of 5 months. A series of six pilot reliability rounds (in alternating pairs) was conducted using lesson videos and transcripts. The coders measured agreement on the CDAS codes and engaged in extended discussions aimed at distinguishing between codes and establishing formal rules for identifying the relevant acts. Definitions of categories underwent extensive trial and refinement to ensure they were meaningful and to minimise inference levels; this process included increasing stringency to apply categories only in explicit cases of, for example, reasoning or reference back (signalled by linguistic markers in each case). This process included specifying expected key words and sentence stems; reasoning in particular was only coded when one of the key words was present, although this alone was insufficient to assume this interactional function; the discourse context was also taken into account. Likewise, Hardman (2019) took discourse markers (e.g. "because", "but") and signaling words (e.g. "reason", "I think") into account when resolving ambiguities and differences between coders undertaking reliability testing.

A document containing correct and incorrect illustrative examples of each code was produced. Once the scheme had been finalised and inconsistencies ironed out, a final reliability exercise included 15% of the project's first year sample of lessons. Final reliability figures are depicted against each CDAS category in Appendix C. The range of Cohen's kappa for moves was 0.58-0.80 (mean 0.68). These figures are very similar to those for SEDA clusters¹³, despite the larger number of categories in CDAS (12 versus 8), so our refinements had been helpful in this sense (although moving from utterance to turn level coding must have helped too). Reasoning (RE) was the most reliable category and salience of reasoning may partly explain its prominence in the dialogue literature. Nevertheless, reliability figures are far from perfect, even for these fairly broad

¹² It was notable, however, that there was considerable disagreement between the experts about these levels within transcripts perceived as falling in the middle of the dialogicality spectrum. While there is approximate consensus in the field concerning the key features of dialogue encapsulated in our scheme, how these are effectively operationalised seems to vary in interpretation.

¹³ The range for SEDA clusters was 0.54-0.88 and mean was 0.70: see Hennessy et al. (2016) for more details of reliability figures.

categories, and this reflects the residual levels of inference that inevitably persist in function-based coding. We note that Huang et al. (2019) achieved higher levels of reliability with SEDA in a Chinese context where traditional pedagogy meant that turns were very short. It is ironic that the more dialogic the interaction and extended the turns in a given context, the less reliable the researchers' coding is likely to be.

Reliability of rating scales was iteratively examined in parallel, but this exercise was far more straightforward since there are only five categories and three possible levels. Reliability is higher for these scales: see Appendix D. The range for rating scales was 75-92% agreement (mean 83%).

5. Conclusions and future directions

This article has considered some of the thorny methodological issues that face researchers attempting to handle “the complexity of doing credible, compelling, rigorous classroom discourse analysis” (Park et al., 2017, p.19). It offers some insights and decision-making strategies that may prove helpful in developing and adapting schemes for micro-level coding and in considering complementary measures. The process was illustrated with an authentic example of adapting SEDA, a comprehensive scheme that had been developed over 3 years by a large international team. It had been carefully mapped against a wide range of theoretical perspectives in the literature and designed to be applied across diverse sociocultural contexts, thereby lending itself to further repurposing. Its fine-grained nature made it likewise suitable for condensing and reformulating.

The article presented two valid and reliable, theory-informed turn-level coding and lesson-level rating tools (together comprising CDAS) that can potentially now also be tested across a wide range of educational contexts. In particular they could support further exploration of the kinds of classroom interaction and student participation that are – and are not – perceived to be associated with student learning. These new instruments have already begun to be taken up and, like the original SEDA, may prove useful for other researchers in the field. Oliver (2018) has modified the CDAS scheme to capture reasoning styles specific to English: genre-, analogy- and language-based reasoning. We ourselves are collaborating with colleagues in Hong Kong to incorporate knowledge building and inquiry categories and examples into a coding scheme for technology-mediated dialogue. Building on such tools allows researchers to shortcut the initial development process and consider how best to modify them to address new needs arising in new contexts. Overall, then, this work makes a significant methodological contribution to the field of dialogue research.

Our large-scale study was clearly heavily dependent on systematic coding followed by quantification. CDAS served these purposes very well in answering our main research questions. Nevertheless, its categories are fairly broad-grained and some of the tantalising findings emerging called for further in-depth, qualitative scrutiny to probe more deeply into which specific forms of dialogue – *within* the original coding categories – are most productive. Our recent exploration of this kind was conducted in conjunction with concordance analysis and focused on a small sample of high performing classes with a dialogic ethos (Hennessy et al., 2019). It has yielded some further insights into the nature of the three markers associated with learning gains (Howe et al., 2019).

The work also carries practical implications for teacher development. The Teacher-SEDA (<http://bit.ly/T-SEDA>) comprises a substantial resource pack that supports practitioners in

understanding and developing dialogic approaches through designing and leading reflective inquiries into their own practices. Tools provided include the student participation rating scale and a coding scheme based heavily on CDAS, with definitions and examples for categories formulated in accessible language, and adaptable for different purposes and contexts; accompanying time sampling tools for live and recorded observations are available. Unusually, Teacher-SEDA supports practitioner-researchers in their own systematic analysis, charting change over time and measuring impacts on students' learning and participation. A recent trial involved over 200 practitioners in schools and tertiary education institutions in eight countries conducting dialogue-focused inquiries leading to documented impacts on practice and thinking; there are versions of the resource pack openly available in English, Spanish and Chinese. The notion of “classroom dialogue” has thereby expanded to span pre-school to postgraduate student levels of education, with application of the coding scheme perceived as the pivotal stimulus for change.

An exponential increase in research conducted in the field of classroom interaction over the last two decades has been charted (Song, Chen, Hao, Liu, & Lan, 2019). The advent of ever-improving machine learning techniques (e.g. Kelly, Olney, Donnelly, Nystrand, & D'Mello, 2018) and wearable interaction sensors (Rosen, Dietz, & Kazi, 2018), plus the evolving expansion of coding tools to capture multimodal interactions (Hennessy, 2020b; Norris, 2004), indicate some of the promising new horizons in the field. The time is ripe for researchers, practitioners and professional development leaders to explore creative and complementary new approaches for analysing and developing dialogue in classrooms, both micro-level coding and other approaches. We hope that this article helps to inform their work.

References

- Alexander, R. J. (2008). *Towards dialogic teaching. Rethinking classroom talk* (Fourth ed.). Cambridge: Dialogos UK Ltd, <http://www.robinaalexander.org.uk/dialogos.htm>.
- Alexander, R. J., F. Hardman, J. Hardman, T. Rajab, and M. Longmore (2017). *Changing Talk, Changing Thinking: Interim Report from the In-House Evaluation of the CPRT/UoY Dialogic Teaching Project*. York: University of York. <http://www.robinaalexander.org.uk/wp-content/uploads/2017/07/Alexander-et-al-EEF-in-house-interim-report-final-170714.pdf>.
- Amodia-Bidakowska, A. and Hennessy, S. (2018, Sept.) *Disciplinary Dialogues: examining the influence of subject cultures and activity type on classroom dialogue in English primary schools*. Paper presented at the Annual Conference of the British Educational Research Association (BERA), Newcastle.
- Asterhan, C.S.C. & Schwarz, B.B. (2016) Argumentation for learning: Well-trodden paths and unexplored territories, *Educational Psychologist*, 51:2, 164-187, DOI: 10.1080/00461520.2016.1155458
- Bakhtin, M. M. (1981). *The Dialogic Imagination*. Austin, TX: University of Texas Press.
- Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29(4), 399-410.
- Boyd, M. P., & Markarian, W. C. (2015). Dialogic teaching and dialogic stance: moving beyond interactional form. *Research in the Teaching of English*, 49(3), 272-296.
- Brauner, E. (2018). Coding interaction. (pp. 165-190). In E. Brauner, M. Boos. and M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis*. Cambridge: Cambridge University Press.
- Brauner, E., Boos, M. and Kolbe, M. (Eds.) (2018). *The Cambridge handbook of group interaction analysis*. Cambridge: Cambridge University Press.
- Chinn, C.A., Anderson, R.C., & Waggoner, M.A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, 36, 378-411. doi:10.1598/rrq.36.4.3
- Clarke, D. (2001) (Ed.). *Perspectives on practice and meaning in mathematics and science classroom*. Dordrecht: Kluwer Academic.
- Correnti, R., Stein, M. K., Smith, M., Scherrer, J., McKeown, M., Greeno, J., et al. (2015). Improving teaching at scale: Design for the scientific measurement and learning of discourse practice. In L. B. Resnick, C. S. C. Asterhan & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington DC: American Educational Research Association Rowman & Littlefield.
- Eagan, B.R., Rogers, R., Serlin, R., Ruis, A.R., Irgens, G.A., & Shaffer, D.W. (2017). *Can we rely on IRR? Testing the assumptions of inter-rater reliability*. Paper presented at International Conference on Computer Supported Collaborative Learning, Philadelphia, PA.
- Edwards, A. D., & Westgate, D. (1994). *Investigating classroom talk* (2nd ed.). London: Falmer Press.
- Elbers, E. (1996). Cooperation and social context in adult-child interaction. *Learning and Instruction*, 6(4), 281-286. [http://dx.doi.org/10.1016/s0959-4752\(96\)00016-](http://dx.doi.org/10.1016/s0959-4752(96)00016-)
- Elizabeth, T., Ross Anderson, T. L., Snow, E. H. and Selman, R. L. (2012). Academic discussions: An analysis of instructional discourse and an argument for an integrative assessment

- framework. *American Educational Research Journal* 49 (6), 1214-1250. doi:10.3102/0002831212456066.
- Fisher, R. (2007). Dialogic teaching: Developing thinking and metacognition through philosophical discussion. *Early Child Development and Care*, 177(6-7), 615–631. <http://dx.doi.org/10.1080/03004430701378985>.
- Freire, P. (1996). *Pedagogy of the oppressed*. London: Penguin.
- Garfinkel, H. (1974). The origins of the term ‘ethnomethodology.’ In *Ethnomethodology: Selected readings* (pp. 15–18). Harmondsworth: Penguin.
- Gee, J. P., & Green, J. L. (1998). Discourse analysis, learning, and social practice: A methodological study. In *Review of Research in Education* (Vol. 23, pp. 119-169): American Educational Research Association.
- Glüer, M. (2018). Software for coding and analyzing interaction processes (pp.245-274). In E. Brauner, M. Boos. and M. Kolbe (Eds.), *The Cambridge Handbook of Group Interaction Analysis*. Cambridge: Cambridge University Press.
- Greeno, J. G. (2016). Classroom talk sequences and learning. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socialising intelligence through academic talk and dialogue* (pp. 255–262). Washington, DC: American Educational Research Association.
- Hardman, J. (2019). Analysing student talk moves in whole class teaching. In N. Mercer, R. Wegerif & L. Major (Eds.), *International Handbook on Dialogic Education*. London: Routledge.
- Hennessy, S. (2011). The role of digital artefacts on the interactive whiteboard in mediating dialogic teaching and learning. *Journal of Computer Assisted Learning*, 27(6), 463-586. doi:10.1111/j.1365-2729.2011.00416.x
- Hennessy, S. (2020a). Analytical coding schemes for classroom dialogue. In R. Kershner, S. Hennessy, R. Wegerif & A. Ahmed (Eds.), *Research methods for educational dialogue*. London: Bloomsbury Academic.
- Hennessy, S. (2020b). Methods for researching technology-mediated dialogue. In R. Kershner, S. Hennessy, R. Wegerif & A. Ahmed (Eds.), *Research methods for educational dialogue*. London: Bloomsbury Academic.
- Hennessy, S., Mercer, N. Calcagni, E., Leung, A. & Lim, J. (2019). *What forms of teacher-student dialogue and student participation are productive for learning?* In symposium “Classroom talk that promotes student learning, motivation, and communication” convened by C. van der Veen & A. Wischgoll. Biennial Conference of the European Association for Research in Learning and Instruction (EARLI), Aachen, Germany.
- Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., ... Barrera, M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction* 9, 16-44.
- Hofmann, R., & Ruthven, K. (2018). Operational, interpersonal, discussional and ideational dimensions of classroom norms for dialogic practice in school mathematics. *British Educational Research Journal*, 44(3), 496-514. Retrieved from <https://doi.org/10.1002/berj.3444> doi:10.1002/berj.3444
- Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43, 325–356.

- Howe, C., Hennessy, S., Mercer, N., Vrikki, M. & Wheatley, L. (2019). Teacher-student dialogue during classroom teaching: Does it really impact upon student outcomes? *Journal for the Learning Sciences*. <https://doi.org/10.1080/10508406.2019.1573730>.
- Howe, C., McWilliam, D., & Cross, G. (2005). Chance favours only the prepared mind: incubation and the delayed effects of peer collaboration. *British Journal of Psychology*, *96*, 67-93.
- Howe, C., & Mercer, N. (2007). *Children's social development, peer interaction and classroom learning*. The primary review (research survey 2/1b). Cambridge: University of Cambridge.
- Howe, C., & Mercer, N. (2016). Commentary on the papers. *Language and Education*, *31*(1), 83-92. doi: 10.1080/09500782.2016.1230126
- Howe, C. J., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, *10*, 361–391.
- Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., ... Donaldson, C. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction*, *17*, 549-563. doi:10.1016/j.learninstruc.2007.09004
- Huang, Yang & Li (2019). On the road to participatory pedagogy: A mixed-methods study of pedagogical interaction in Chinese kindergartens. *Teaching and Teacher Education*, *85*, 81-91. Retrieved from <https://doi.org/10.1016/j.tate.2019.06.009>.
- Hymes, D. (1972). Models of interaction in language and social life. In J. J. Gumperz, & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35–71). London: Basil Blackwell. doi:10.1002/9780470758434.ch1
- Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. *Learning and Instruction*, *21*, 379–393.
- Kelly, S., Olney, A. M., Donnelly, P. J., Nystrand, M., & D'Mello, S. K. (2018). Automatically measuring question authenticity in real-world classrooms. *Educational Researcher*, *47* (7), 451-464.
- Kershner, R. (2020). Dialogue, participation and social relationships. In R. Kershner, S. Hennessy, R. Wegerif & A. Ahmed (Eds.), *Research methods for educational dialogue*. London: Bloomsbury Academic.
- Kershner, R., Hennessy, S., Wegerif, R., & Ahmed, A. (2020). *Research methods for educational dialogue*. London: Bloomsbury Academic.
- Kim, M.-Y., & Wilkinson, I. A. (2019). What is dialogic teaching? Constructing, deconstructing, and reconstructing a pedagogy of classroom talk. *Learning, Culture and Social Interaction*, *21*, 70-86.
- Kumpulainen, K., & Lipponen, L. (2010). Productive interaction as agentic participation in dialogic enquiry. Chapter 3. In K. Littleton, & C. Howe (Eds.). *Educational Dialogues: Understanding and Promoting Productive Interaction* (pp. 48–63). Abingdon, Oxon: Routledge.
- Kumpulainen, K., & Wray, D. (Eds.). (2002). *Classroom interaction and social learning: From theory to practice*. London: Routledge.
- Lefstein, A. (2008). Changing classroom practice through the English national literacy strategy: A micro-international perspective. *American Educational Research Journal*, *45*, 701–737.

- Lefstein, A., Snell, J. and Israeli, M. (2015). From moves to sequences: expanding the unit of analysis in the study of classroom discourse. *British Educational Research Journal*. DOI: 10.1002/berj.3164.
- Linell, P. (2001). *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins.
- Ludvigsen, K., Krumsvik, R. J. & Breivik, J. (2020). Behind the scenes: Unpacking student discussion and critical reflection in lectures. *British Journal of Educational Technology*.
- Maine, F., Rojas Drummond, S., Hofmann, R., Barrera, M-J. (2020). Symmetries and asymmetries in children's peer-group reading discussions in Mexico and the UK. *Australian Journal of Language and Literacy*, 43(1), 17–32.
- Matusov, E. (2011). Authorial teaching and learning. In E. J. White & M. Peters (Eds.), *Bakhtinian pedagogy: Opportunities and challenges for research, policy and practice in education across the globe* (pp. 21-46). New York: Peter Lang Publishers.
- Maybin, J. (2006). *Children's voices: Talk, knowledge and identity*. Basingstoke: Palgrave Macmillan.
- Mayer, S. (2018). Studying student efforts to frame (and pursue) academically substantive inquiries. Poster presented at European Association for Research in Learning and Instruction (EARLI) SIG 20/26 *Inquiry Learning and Argumentation, Dialogue and Reasoning Conference*, Jerusalem.
- Mazzei, L. A. (2014). Beyond an easy sense: A diffractive analysis. *Qualitative Inquiry*, 20(6), 742-746. doi:10.1177/1077800414530257
- Mercer, N. (2000). Development through dialogue. In N. Mercer (Ed.), *Words and minds: How we use language to think together* (pp. 131-166). London: Routledge.
- 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, 1-14.
- Mercer, N., & Dawes, L. (2014). The study of talk between teachers and students, from the 1970s until the 2010s. *Oxford Review of Education*, 40(4), 430-445. doi:10.1080/03054985.2014.934087
- Mercer, N., & Dawes, L. (2008). The value of exploratory talk. In N. Mercer & S. Hodgkinson (Eds.), *Exploring talk in school* (pp.55-71). London: Sage.
- Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, 30(3), 367-385.
- Mercer, N., & Littleton, K. (2007). *Dialogue and the Development of Children's Thinking*. London: Routledge.
- Merleau-Ponty, M. (1968). *The visible and the invisible, followed by working notes*, English trans. by A. Lingis. Evanston: Northwestern University Press.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.

- Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools. Professional development approaches for academically productive discussions. In L.B. Resnick, C.S.C. Asterhan & S.N. Clarke (Eds.) *Socializing intelligence through academic talk and dialogue* (pp. 347-362). Washington, DC: American Educational Research Association.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and civic life. *Studies in Philosophy and Education*, 27(4), 283–297.
- Molinari, L., Mameli, C., & Gnisci, A. (2013). 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. doi:10.1111/j.2044-8279.2012.02071.x
- Mortimer, E. F. (2005). *Dialogic and authoritative discourse: A constitutive tension of science classroom*. Universidade Federal de Minas Gerais. ICAR - CNRS - Université Lumière Lyon 2. Retrieved from http://icar.univ-lyon2.fr/ecole_thematique/analyse_video/documents/Dialogic_and_authoritative_discourse.pdf
- Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead: Open University Press.
- Myhill, D. (2006) Talk, talk, talk: teaching and learning in whole class discourse, *Research Papers in Education*, 21:1, 19-41. doi: 10.1080/02671520500445425.
- Norris, S. (2004). *Analyzing multimodal interaction: a methodological framework*. New York and London: Routledge.
- Nystrand, M. (1988). CLASS (Classroom language assessment system) 2.0: A Windows laptop computer system for the inclass analysis of classroom discourse. Madison: Wisconsin Center for Education Research.
- Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York, London: Teachers' College Press.
- Nystrand, M., Wu, A., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, 35 (3), 135–198. doi:10.1207/s15326950dp3502_3
- Oliver, M. (2018). Exploring and developing reasoning in primary English. *Proceedings of the Imagining Better Education Conference*, Durham University, 19 March, 2018.
- Osborne, J. Berson, E., Borko, H., Busch, K.C., Zaccarelli, F.G., Million, S., Tseng, A. (2015). Assessing the quality of classroom discourse in science classrooms. In *16th Biennial Conference of the European Association for Research in Learning and Instruction*, Limassol, Cyprus (751). 25-29 Aug, 2015.
- Park, J. Y., Michaels, S., Affolter, R., & O'Connor, M. C. (2017). Traditions, research and practice supporting academically productive classroom discourse. *Oxford Research Encyclopedia of Education*. Oxford: Oxford University Press.
- Pianta, R. C., & La Paro, K. M. (2003). *Classroom Assessment Scoring System (CLASS): Guide and training manual*. University of Virginia, Charlottesville.
- Resnick, L., Asterhan, C., & Clarke, S. (2015). *Socializing intelligence: Through academic talk and dialogue*. Washington, D.C: AERA.

- Resnick, L. B., & Michaels, S. (2010). How (well structured) talk builds the mind. In D. Preiss, & R. Sternberg (Eds.). *Innovations in educational psychology* (pp. 163–194). New York: NY: Springer.
- Resnick, L. B., Michaels, S., & O'Connor, M.C. (2010). How (well structured) talk builds the mind. In D. Preiss, D., & Sternberg, R. (Eds.), *Innovations in educational psychology* (pp. 163–194). New York, NY: Springer.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford: Oxford University Press.
- Rojas-Drummond, S., Torreblanca, O., Pedraza, H., Vélez, M., and Guzmán, K. (2013). 'Dialogic scaffolding': Enhancing learning and understanding in collaborative contexts. *Learning, Culture and Social Interaction*, 1, 11-21.
- Rosen, Dietz & Kazi, Beyond coding interaction. (pp.142-162). In E. Brauner, M. Boos. and M. Kolbe (Eds.). *The Cambridge handbook of group interaction analysis*. Cambridge: Cambridge University Press.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.
- Saville-Troike, M. (2003). *The ethnography of communication: An introduction* (3rd ed.). Oxford: Blackwell Publishing Ltd.
- Schwarz, B. B. and M. Baker (2017). *Dialogue, argumentation and education: History, theory and practice*. New York, Cambridge University Press.
- Sedova, K., Sedlacek, M., & Svaricek, R. (2016). Teacher professional development as a means of transforming student classroom talk. *Teaching and Teacher Education*, 57, 14-25.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart.
- Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.
- Skidmore, D., & Murakami, K. (2016). *Dialogic pedagogy: The importance of dialogue in teaching and learning*. Bristol: Multilingual Matters.
- Song, Y., Chen, X., Hao, T., Liu, Z., & Lan, Z. (2019). Exploring two decades of research on classroom dialogue by using bibliometric analysis. *Computers & Education*, 137, 12-31. doi: <https://doi.org/10.1016/j.compedu.2019.04.002>
- Soter, A. O., Wilkinson, I. A., Murphy, K. P., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research*, 47(6), 372-391. doi:10.1016/j.ijer.2009.01.001
- St. Pierre, E. A., & Jackson, A. Y. (2014). Qualitative Data Analysis After Coding. *Qualitative Inquiry*, 20(6), 715-719. doi: 10.1177/1077800414532435
- Tschan, F. Zimmerman, J. and Semmer, N.K. (2018) Rules for coding scheme development (pp.191-207). In E. Brauner, M. Boos. and M. Kolbe (Eds.). *The Cambridge handbook of group interaction analysis*. Cambridge: Cambridge University Press.
- van Drie, J., & Dekker, R. (2013). Theoretical triangulation as an approach for revealing the complexity of a classroom discussion. *British Educational Research Journal*, 39(2), 338-360.
- van de Pol, J., Volman, M., & Beishuizen, J. (2011). Patterns of contingent teaching in teacher–student interaction. *Learning and Instruction*, 21(1), 46–57.

- van der Veen, C., de Mey, L., van Kruistum, C., & van Oers, B. (2017). The effect of productive classroom talk and metacommunication on young children's oral communicative competence and subject matter knowledge: An intervention study in early childhood education. *Learning and Instruction, 48*, 14-22. doi:10.1016/j.learninstruc.2016.06.001
- Vrikki, M., Wheatley, L., Howe, C., Hennessy, S. & Mercer, N. (2019). Dialogic practices in primary school classrooms. *Language and Education 33*(1), 85-100.
<https://doi.org/10.1080/09500782.2018.1509988>.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., et al. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research, 63*, 79-93. doi: 10.1016/j.ijer.2013.02.001
- Wegerif, R. (2007). *Dialogic, Education and Technology: Expanding the Space of Learning*. New York: Springer.
- Wegerif, R. (2020). Foundations for research on educational dialogue. In R. Kershner, S. Hennessy, R. Wegerif & A. Ahmed (Eds.), *Research methods for educational dialogue*. London: Bloomsbury Academic.
- Wegerif, R., & Mercer, N. (1997). Using computer-based text analysis to integrate qualitative and quantitative methods in research on collaborative learning. *Language and Education, 11*(4), 271-286. Retrieved from <https://doi.org/10.1080/09500789708666733>.
- Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. Cambridge: Cambridge University Press.
- Wells, G. (2009). *The Meaning Makers: Learning to Talk and Talking to Learn* (2nd ed.). Bristol: Multilingual Matters.
- Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *Journal of the Learning Sciences, 15*(3), 379-428.
- Wells, G. with DICEP (2001). *Coding scheme for the analysis of classroom discourse*. Retrieved from https://people.ucsc.edu/~gwells/Files/Courses_Folder/documents/CodingManual.pdf.
- Wilkinson, I. A., Soter, A. O. & Murphy, K. P. (2010) Developing a model of quality talk about literary text, in: M. G. McKeown & L. Kucan (Eds) *Bringing reading researchers to life: Essays in honor of Isabel L. Beck* (pp. 142–169). New York: Guilford Press.
- Wilkinson, I.A.G., Reznitskaya, A., Bourdage, K., Oyler, J., Glina, M., Drewry, R., ... Nelson, K. (2017). Toward a more dialogic pedagogy: Changing teachers' beliefs and practices through professional development in language arts classrooms. *Language and Education, 31*, 65-82. doi:10.1080/09500782.2016.1230129.
- Wood, D. J., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Acknowledgements

Development of CDAS was undertaken as part of the project “Classroom Dialogue: Does it really make a difference for student learning?”, led by Howe, Hennessy and Mercer (2015-2017): <http://tinyurl.com/ESRCdialogue>. The work was funded by the Economic and Social Research Council under grant ES/M007103/1. We are grateful to our collaborators and coding team members, including Annabel Amodia-Bidakowski, Elisa Calcagni and Lisa Wheatley. Development of SEDA was carried out in conjunction with Mexican colleagues led by Sylvia Rojas-Drummond for a project (<http://tinyurl.com/BAdialogue>) funded by the British Academy (2013-2015).

Appendix A. Condensed version of SEDA ©2016: Cluster and Code Summary

I – Invite elaboration or reasoning

	Ask for explanation or justification of another’s contribution
I2	Invite building on / elaboration / (dis)agreement / evaluation of another’s contribution or view
I3	Invite possibility thinking based on another’s contribution
I4	Ask for explanation or justification
I5	Invite possibility thinking or prediction
I6	Ask for elaboration or clarification

P – Positioning and Coordination

P1	Synthesise ideas
P2	Evaluate alternative views
P3	Propose resolution
P4	Acknowledge shift of position
P5	Challenge viewpoint
P6	State (dis)agreement/ position

RD – Reflect on dialogue or activity

RD1	Talk about talk
RD2	Reflect on learning process/ purpose/ value/ outcome
RD3	Invite reflection about process/ purpose/ value/ outcome of learning

E – Express or invite ideas

E1	Invite opinions/beliefs/ ideas
E2	Make other relevant contribution

R – Make reasoning explicit

R1	Explain or justify another’s contribution
R2	Explain or justify own contribution
R3	Speculate or predict on the basis of another’s contribution
R4	Speculate or predict

B – Build on ideas

B1	Build on /clarify others' contributions
B2	Clarify/elaborate own contribution

C – Connect

C1	Refer back
C2	Make learning trajectory explicit
C3	Link learning to wider contexts
C4	Invite inquiry beyond the lesson

G – Guide direction of dialogue or activity

G1	Encourage student-student dialogue
G2	Propose action or inquiry activity
G3	Introduce authoritative perspective
G4	Provide informative feedback
G5	Focusing

Appendix B. Adaptation of SEDA for CDAS and rationale for changes

Consideration	SEDA	Rationale and strategy for change	CDAS
Scope: Treatment of non-dialogic discourse	Optional <i>Expression of ideas</i> cluster for statements/ propositions and invitations; other non-dialogic talk <i>Uncoded</i>	Adaptation needed to retain core principles of productive educational dialogue, whilst capturing contributions that sustain talk and enable measurement of ratio of dialogic to non-dialogic interaction (adding <i>Other invitations</i> to <i>Uncoded</i> : Howe et al., 2019).	Category of non-dialogic or <i>Other invitations</i> (aside from ELI, REI, CI or Invite RB), e.g. closed questions Unconnected student expressions of ideas or views and other non-dialogic talk was " <i>Uncoded</i> "
Broader granularity of scheme desired	33 codes (including optional <i>Express ideas</i> codes) in 8 clusters	Need to reduce number of options for coders of large dataset and achieve acceptable inter-coder agreement levels, yet be representative of field as whole. We aimed for minimum scheme to capture all key moves. Desire to retain some finer distinctions in SEDA, e.g. reference back to prior contributions versus reference to wider context outside classroom.	11 dialogue move codes (plus <i>Other Invitations</i>)
Mutually exclusivity within clusters	Not exclusive – hierarchical. Where 2 codes within a cluster are applicable, the more dialogic code is applied so there is only one assigned per cluster.	Frequency counting necessitated eliminating overlap between SEDA categories to make them mutually exclusive and avoid double counting. Codes in Positioning and Coordination cluster proved slightly problematic as framed: the description for (Dis)agreement states that "positioning in relation to other must be explicit", but in practice this must occur via elaboration or reasoning, leading to some double counting. Rules were defined to prioritise codes where more than one applied within a turn or even utterance, but emerging codes are at least conceptually exclusive.	Exclusive, except that reasoning and co-ordination/synthesis each occur respectively in some form in definitions of a pair of codes (RE/RC and SC/RC) although these are treated as exclusive when coding, since RC is prioritised over both SC and RE. ¹⁴
Finer granularity of category / more	<i>Positioning & coordination</i> combined in one cluster.	Ideas can be coordinated with/without preference so <i>coordination</i> construct separated out from <i>positioning</i> . Need to address concerns of some theorists in the argumentation tradition who stress that reasoned	More refinement of coordination categories to capture invitations to relate to ideas in order to summarise, compare or critically evaluate them (<i>Invite coordination</i> : CI) and distinguishing <i>Simple</i>

¹⁴ Coordination is so rare that frequency counts are not affected in practice.

stringency desired	Counter-argument coded via <i>state (dis)agreement/position + reasoning</i> combination. Reasoned coordination coded via <i>synthesise ideas + reasoning</i>	coordination is important for learning, albeit rare in untrained students. Finer distinctions concerning synthesis of ideas and reasoning were introduced via several “coordination of ideas” codes, where bringing in additional information or evidence is key.	<i>Coordination (SC)</i> from <i>Reasoned Coordination (RC)</i> in capturing responses to such an invitation. RC is more stringent in that it evaluates alternative perspectives with reasoning and draws on evidence, theory or a mechanism for justification. SC simply synthesises or summarises collective ideas, or it compares different perspectives.
Broader granularity of category desired (links with validity and reliability)	Separate codes for speculation/ prediction/ possibility thinking (Invite / Express clusters) and justification not required	Emphasis on justification and accountability to knowledge within literature (validity) Justification/prediction codes were collapsed into a broader <i>reasoning</i> category to reduce codes and tighten reasoning threshold.	Prediction integrated into <i>reasoning</i> code and requires justification
	<i>Refer back, Make learning trajectory explicit, Link learning to wider contexts</i> and <i>Invite inquiry beyond the lesson</i> are all separate codes. Challenge is embodied in several codes: <i>Challenge viewpoint; State (dis)agreement/ position; Invite (dis)agreement with another’s contribution or view</i>	In the interests of reduction several categories collapsed SEDA codes where they were evidently related	2 codes: <i>Reference back (RB)</i> and <i>Reference to wider context (RW)</i> , including invitations as well as statements; RB included <i>Make learning trajectory explicit</i> and RW includes <i>Invite inquiry beyond the lesson</i> Single code: <i>Querying</i> (Doubting, full/partial disagreement, challenging or rejecting a statement): includes simple No response
	Distinction of own vs. other’s contribution under reasoning and elaboration	Joint ownership of ideas means that distinguishing individual contributions is often hard; codes were collapsed to increase reliability .	No distinction between own and other’s contribution

<p>Broader granularity of category definition / less stringency desired</p>	<p>Agreement required at least two positions to have been expressed previously so that one is chosen over the other.</p>	<p>SEDA formulation is very dialogic but potentially overlaps with reasoning and elaboration categories. Categories expanded (as in cell to right).</p>	<p><i>Agreement (A)</i> and <i>Querying (Q)</i> categories capture the extent to which ideas are being accepted or challenged/ questioned. A includes repetition or simple agreement such as “yes” or “good” and thus does not constitute an optimal form of dialogue in itself although in combination with <i>Elaboration</i> or <i>Invitations for elaboration</i>, it represents Nystrand et al.’s (2003) high level evaluation move. This category proved not very useful and was later abandoned. Likewise, Q includes simple “No” statements which in retrospect was rather too broad a definition (given that a small number of teachers’ use of “No” on its own served to cut off the dialogue rather than to challenge students productively).</p>
--	--	---	--

Appendix D: Rating Scales

	0 Not evident	1 Teacher-led	2 Teacher-led with student involvement
Aims and objectives	Aims (goals, objectives, plans etc) are not made explicit	Teacher “announces” lesson aims, with at most non-interactive, e.g. written, responses from students	Aims (goals, objectives, plans etc) are to some extent negotiated with students or by students themselves (<i>does not include negotiation of dialogue rules – that is measured in “Focusing on dialogue” scale</i>)
Monitoring and guidance	Monitoring and guidance are not apparent	Teacher watches while students carry tasks out and offers suggestions and/or evaluations	Teacher offers help without taking over. Monitoring and guidance is to some extent tailored through dialogue towards students’ needs (i.e. scaffolding or focusing, e.g. teacher strategically balances dialogic and authoritative)
Reflection on learning process	Reflection on the learning process or trajectory does not take place	Teacher reports on what has been achieved or makes learning process / trajectory explicit, perhaps inviting reflection	Reflection on learning process or trajectory is discussed with students or by students themselves. This reflection can include a joint understanding of what has been achieved, evaluation of how activities have gone, understanding of where learning is headed. Includes students’ own metacognitive reflection on learning process itself.
Focusing on dialogue	No explicit focus on dialogue is apparent	Teacher introduces, models or reminds students of target dialogic practices, e.g. ground rules to be followed, inclusive turn taking.	Teacher and students or students themselves negotiate target dialogic practices, eg ground rules, perhaps along with reminders / modelling. It may also include students being given or taking responsibility for managing the dialogue, as well as students being involved in evaluating effectiveness of dialogic practices.
Student participation	Public exchanges consist in questioning and succinct students’ contributions or students don’t have opportunities to discuss their ideas publicly	Students express their ideas publicly at length in whole-class situation and group work but they don’t engage with each other’s ideas	Multiple students express their ideas publicly at length in whole-class situation and group work (consider what is audible in the video) AND In doing so, they engage with each other’s ideas , for example by referring back to their contributions, challenging or elaborating on them (e.g. “Sam had such a great idea, look [demonstrates]”). This includes spontaneous or teacher-prompted participation.

Appendix C: Summary of CDAS

CODE	DEFINITION	KAPPA
Elaboration invitations (ELI)	Invites building on, elaboration, evaluation, clarification of own or another's contribution. E.g. "Have you noticed anything else that the poet uses?", "I agree with you it makes a strong picture, but what do you picture?"	.62
Elaboration (EL)	Builds on, elaborates, evaluates, clarifies own or other's contribution. E.g. [In reply to "It's sort of describing how you do it"] "Yes, it's got a good emphasis and a good use of vocabulary"	.63
Reasoning invitations (REI)	Explicitly invites explanation/justification of a contribution or speculation (new scenarios) /prediction/hypothesis. E.g. "Why do you think the bottle floats?", "If Billy's correct, what would you expect to happen?"	.73
Reasoning (RE)	Provides an explanation or justification of own or another's contribution, or speculates, predicts, hypothesizes with grounds given. E.g. [After "He came back"] "because he made a promise".	.80
Co-ordination invitations (CI)	Invites synthesis, summary, comparison, evaluation or resolution based on two or more contributions. E.g. "Would anyone like to summarize the ideas we've been hearing?"	NC
Simple co-ordination (SC)	Synthesises or summarises collective ideas (including own and/or others' ideas). Compares, resolves or evaluates different opinions, perspectives and beliefs. E.g. "Some of you are talking about weight and some are talking about size; both matter – things float when they're light for their size".	.76
Reasoned co-ordination (RC)	Compares, evaluates, resolves two or more contributions in a reasoned fashion. It includes all SC descriptors plus a counter-argument, reasoned rebuttal, two partial truths, e.g. drawing on evidence, theory or a mechanism. E.g. "We've been arguing about how much of personality is inherited; twin studies show conclusively it's 50%".	NC
Agreement (A)	Explicit acceptance of or agreement with a statement(s). E.g. "Brilliant", "Good", "Yeah", "Okay", "I agree with X...".	.69
Querying (Q)	Doubting, full/partial disagreement, challenging or rejecting a statement. Includes a simple "no" response when it shows rejection of an idea; not when in response to a question. E.g. "Do you really think these angles are the same?"	.62
Reference back (RB)	Introduces reference to previous knowledge, beliefs, experiences or contributions (includes procedural references) that are common to the current conversation participants. Includes <i>inviting</i> reference back. E.g. "Can anyone remember which of the animals we saw at the zoo are nocturnal?"	.62
Reference to wider context (RW)	Making links between what is being learned and a wider context by introducing knowledge, beliefs, experiences or contributions from outside of the subject being taught, classroom or school. Includes <i>inviting</i> reference to wider context. E.g. "It's like in Macbeth where the storm builds into it".	.58
Other invitations (OI)	Invitations for verbal contributions (e.g. opinions, ideas, beliefs), except for those coded as ELI, REI, CI, RB or RW. E.g. dichotomous questions, student nominations, procedural questions, closed questions ("What is 6 x 2?")	.72