This is the author accepted version, accepted for publication in Learning, Culture, and Social Interaction. Readers should cite the published version available via the journal website
https://doi.org/10.1016/j.lcsi.2017.11.003

A Discursive Approach to the Analysis of Epistemic Cognition

Simon Knight[a], Karen Littleton[b]

a. Simon.knight@uts.edu.au, Faculty of Transdisciplinary Innovation, University of Technology Sydney, Australia
b. Faculty of Wellbeing, Education and Language Studies, The Open University, UK.

A core concern in learning is coming to understand the ways in which claims of knowledge are made. The epistemic cognition literature typically characterises this learning in terms of how learners cognitively conceptualise the source and nature of knowledge. Recent work has offered alternative accounts of epistemic cognition that recognise the discursive nature of the construct. These accounts are derived from analysis of the ways that learners talk about knowledge in tasks such as evaluating scientific claims from sources of varying qualities. In this paper we draw on this recent work to advance a novel approach to the analysis of discourse data in epistemic contexts. This approach is exemplified through its application to an existing dataset, demonstrating both the application of the approach and the particular kinds of discourse that learners engaged in. This discursive approach has the potential for broad application in the learning sciences' treatment of epistemic cognition.

Keywords: Epistemic cognition; epistemological beliefs; discursive psychology; sociocultural theory; sociocultural discourse analysis; dialogue; collaborative learning

1. Introduction

A core concern in learning is coming to understand the ways in which claims of knowledge are made. In developing their understanding of the social and cultural practices through which knowledge claims are made, people become members of particular discourse communities, able to take on board and critique the claims that they encounter. Accounting for the developmental and cognitive processes through which students come to understand knowledge claims has been focal for the area of 'epistemic cognition' within the learning sciences. Building on our recent *socialised* account of epistemic cognition (Knight & Littleton, 2017), in this paper we foreground the *discursive* features of epistemic cognition. In so doing our aim is to exemplify features of our socialised account, in order to develop a generalised approach to understanding particularly *epistemic* types of dialogue. We do this by drawing on empirical work drawn from two rather different datasets, exemplifying the potential of the analytic approach taken to understand this data in light of epistemic cognition.

Our recent paper (Knight & Littleton, 2017) advanced an account of epistemic cognition that builds on social epistemology, to re-specify epistemic cognition in sociocultural terms focusing on the ways that knowledge claims are:

1. Embedded in situated, occasioned communicative practice; i.e., that knowledge claims communicate meaning, and are co-constructed in, social interaction.
2. Pragmatic; i.e., that the communicated meaning is oriented towards some ends, we make claims in *doing* things.
3. Normative; i.e., that the ways in which we make and interact around claims are defined by sets of normative practices.

Discursive psychology (Edwards, 2005; Potter & Edwards, 1999) has often 're-specified' psychological constructs through considering the constructs with respect to language-in-use. In this way, epistemic cognition can be conceived in terms of the ways that its characteristics are implicated in the dialogue

evoked within epistemically salient contexts. In developing our account we follow, for example, Greeno, Collins and Resnick (1992) in highlighting the common history shared by sociocultural and discursive approaches in learning, and emerging pragmatic and social epistemological perspectives in philosophy. A socialised account of epistemic cognition, drawing from social epistemology, thus aligns with sociocultural perspectives on learning, bringing both the individual and collective into view. This shift motivates a refocusing on the ways in which epistemic cognition emerges from, and is co-constructed in, the dialogue implicated in epistemically salient work. Commensurately, this shift re-emphasises language-in-action, alongside or in place of self-report methods (for example, interview, talk aloud, surveys) that have been the focus of much epistemic cognition research to date. Epistemic cognition should thus be understood with respect to language-in-action in the context of collective (such as, but not limited to, small group, class) activities.

The theoretical stance we outline is not intended to replace, but rather to – using the discursive psychology term – *re-specify* cognitive constructs in terms of dialogue. It thus draws on the broad agreement across a set of models of epistemic cognition, on two core features of the construct: What knowledge is, and how one comes to know:

> *There are two dimensions within the first area (knowledge):*
>
> - *Certainty of knowledge: the degree to which knowledge is conceived as stable or changing, ranging from absolute to tentative and evolving knowledge;*
> - *Simplicity of knowledge: the degree to which knowledge is conceived as compartmentalized or interrelated, ranging from knowledge as made up of discrete and simple facts to knowledge as complex and comprising interrelated concepts.*
>
> *There are also two dimensions which can be identified within the second area (knowing):*
>
> - *Source of knowledge: the relationship between knower and known, ranging from the belief that knowledge resides outside the self and is transmitted, to the belief that it is constructed by the self;*
> - *The justification for knowing: what makes a sufficient knowledge claim, ranging from the belief in observation or authority as sources, to the belief in the use of rules of inquiry and evaluation of expertise.*
>
> (Mason, Boldrin, & Ariasi, 2010, p. 69)

In this paper, we outline the communicative and pragmatic nature of the *epistemic* form of dialogue we have observed in a particular set of information seeking tasks in which questions of certainty, simplicity, sourcing, and justification of claims are salient. We do this through presenting a new, secondary analysis of dialogue in a small-group secondary school classroom activity (reported in Knight, Arastoopour, Williamson Shaffer, Buckingham Shum, & Littleton, 2014; Knight & Mercer, 2015, 2016), and dyadic chat data from a university-level task facilitated by a browser add-on and chat tool (reported in Knight, 2016). This analysis:

1. Exemplifies the application of the account.
2. Describes the application of a method for such analysis.
3. Exemplifies that method, drawing attention to the insights provided.

In the following sections we: (1) provide an overview of our social account of epistemic cognition, highlighting its distinctively socio-cultural nature; (2) provide a description of the discursive features of

this account; (3) use this account to motivate a novel approach to the development of a typology for the analysis of epistemic dialogue, that we exemplify through its application to the two sample datasets; (4) conclude, highlighting the key contribution of a theoretically motivated approach to the analysis of epistemic cognition as a discursive feature of learning.

2. A discursive account of epistemic cognition

Language-in-action offers a particular lens onto human epistemic behaviour in contexts that are both constituted in, and by, that language. This view differs from a typical cognitive perspective in that, "stance in discourse is not the transparent linguistic packaging of 'internal states' of knowledge, but rather emerges from dialogic interaction" (Kärkkäinen, 2006, p. 699). Moreover, these stances are normatively situated, that it is taken that "[t]o 'know' is to operate discursively within such a community [of practice]" (Mercer & Howe, 2012, p. 19). In contexts in which there is an epistemic aim to learn or build knowledge, commonality – driven by language – is key; as Edwards and Mercer term it, 'common knowledge' (Edwards & Mercer, 1987). Fundamentally, the social epistemological position motivates a greater attention to understanding knowledge as communicative, recognising that epistemic stance emerges in the interaction between speakers (Kärkkäinen, 2006). Given the particular context of learning, it is important to note van Dijk's (2014) point that text and talk provide resources for that learning, both representing, and constituted through discourse. The development of 'common knowledge' – a shared perspective – is key for the claims that one makes to be understood within the social context.

In Kärkkäinen's work the notion of 'epistemic stance' denotes: "marking the degree of commitment to what one is saying, or marking attitudes toward knowledge. This definition also includes evidential distinctions, or how knowledge was obtained and what kind of evidence the speaker provides for it" (Kärkkäinen, 2006, p. 705); such markers include 'I think', 'he' or 'she' said, 'I don't know' 'I guess', 'I thought', epistemic adverbs such as 'maybe', 'probably', 'apparently', 'of course', and epistemic modal auxiliaries such as 'would', 'must', 'might', 'could', 'will', 'may'. Stance, then, is seen as an emergent, dynamic, and intersubjective property of pragmatic action within particular, social, contexts.

This is not, of course, to suggest that the extant literature on epistemic cognition has ignored such social considerations. How students engage with the resources around them – including in their dialogue with other actors, and the written texts that they encounter – is an important, and social, issue within the existing work. However, the consideration of these features as dynamic and fundamental to the situated emergent properties of an epistemic stance, has not typically been investigated.

2.1 The Learning Context of a Socialised Account of Epistemic Cognition

We have, then, proposed a shift in our understanding of epistemic cognition, with a particular focus on the interactive communicative nature of knowledge claims aimed at pragmatic ends within normative contexts. A body of work, which we discuss further in this section, has analysed dialogue in the context of epistemic cognition. However, these analyses have tended not to focus on language as a tool for thinking in its own right, instead focusing on it as a means to an ends.

Perhaps closest to our own position is that of Österholm, who calls specifically for a discursive psychology approach to describe: "the activity, the discourse, as the site where epistemological beliefs come to existence, through explicit or implicit references to prior experiences (epistemological resources)" (Österholm, 2009, p. 262) and aligns this view with Hammer and Elby's 'resources' model of epistemic cognition (Hammer & Elby, 2003). In a science classroom-based study by one of those authors, Hutchinson and Hammer (2010) do illustrate epistemic perspectives through the use of dialogue excerpts, indicating a distinction between 'absolutist' perspectives on knowledge, and dialogic or

sensemaking dialogue – which we would characterise as being exploratory in nature (described in Wegerif, 2006) – in which ideas are built upon and become threads (or cohesive ties) through a dialogue, with similar claims made in other work (Rosenberg, Hammer, & Phelan, 2006)

We have argued that such dialogue, and its analysis, parallels the kind of dialogue that for example Reznitskaya and Gregory (2013) discuss in relation to conceptual change and epistemic cognition. These authors draw attention to dialogic talk in which evidence is explored and reasons justified, as opposed to 'monologism' in which dialogue aims at a singular truth. They highlight (p.118-119) that 'evaluativist' perspectives are more dialogic in nature, flagging that dialogic contexts:

1. Recognise expertise, and its limits.
2. Centre on ill-structured questions around which one might hold different perspectives.
3. Have a metacognitive character in which students must consider other's perspectives and build awareness of their own position in relation to others.

The kind of dialogic language described by Reznitskaya and Gregory has strong association to a kind of dialogue that we would describe as 'exploratory' in nature. Reznitskaya and Gregory (2013) suggest that epistemic cognition of an 'evaluativist' kind – which they hold to be more sophisticated – is closely associated with the kind of exploratory talk associated with educational gains (Mercer, Dawes, Wegerif, & Sams, 2004; Mercer & Sams, 2006; Mercer, Wegerif, & Dawes, 1999; Rojas-Drummond, Littleton, Hernández, & Zúñiga, 2010). Exploratory dialogue, in contrast to cumulative and disputational dialogue, involves talk in which:

> *Partners engage critically but constructively with each other's ideas. Statements and suggestions are offered for joint consideration. These may be challenged and counter-challenged, but challenges are justified and alternative hypotheses are offered. Partners all actively participate, and opinions are sought and considered before decisions are jointly made. Compared with the other two types, in exploratory talk knowledge is made more publicly accountable and reasoning is more visible in the talk. (Mercer & Littleton, 2007, pp. 58–59)*

Our claim, then, is that through such dialogue students engage in building 'common knowledge' for learning, by using language to think together and build on each other's ideas (for example, Littleton & Mercer, 2013). This type of dialogue emphasises learning and listening from others, expressing *and* explaining ideas, and the use of argument that would be considered appropriate within the practices of the community. In Kärkkäinen's terms, through dialogue, students can build common knowledge and develop an emergent stance that is oriented both towards the historic common knowledge, and to the dynamic and emergent context of the activity. As such, analysis of dialogue offers us insight into the ways in which these epistemic stances are co-constructed within particular *learning* contexts.

3. The Development of an Analytic Approach

3.1 Context of the Datasets

The theoretical context outlined above (and presented in more detail in, Knight & Littleton, 2017) has informed, and been informed by, our developing work on the analysis of epistemic dialogue over a range of learning tasks. Our particular interest has been in the ways that epistemic cognition is

implicated in tasks for which information seeking is required. In other work we have described how such tasks might be orchestrated for the probing of epistemic features of information seeking (Knight, Rienties, Littleton, Tempelaar, et al., 2017), demonstrating that such task design elicits clear epistemic features in student behaviour through the selection of varied queries, websites, and content in written artefacts (Knight, Rienties, Littleton, Mitsui, et al., 2017; Knight, Allen, Littleton, Rienties, & Tempelaar, 2016), and in the dialogue of school-children (reported in, Knight et al., 2014; Knight & Mercer, 2015, and, 2016).

Extant work on dialogue and epistemic cognition has applied a range of approaches to understanding that dialogue. However, no overarching approach or method for the treatment of epistemic dialogue has been described, nor a systematic method for understanding such dialogue across contexts. Despite this, many theoretical perspectives – including but not limited to the socialised account in (Knight & Littleton, 2017) – would recognize the importance of dialogue in expressing, building, and analysing epistemic cognition. This paper addresses this gap, providing an analytic approach to the investigation of epistemic features of dialogue. This approach is exemplified with two datasets that arise from quite different contexts (described below); these differences highlight the utility of the approach taken across diverse settings, and its potential for wider application.

3.2 An Analytic Approach to Epistemic Dialogue

In the following sections we describe an analytic approach for understanding epistemic dialogue, exemplifying its application through discussion of data taken from two studies. The first of these studies (reported in, Knight et al., 2014; Knight & Mercer, 2015, and, 2016) was conducted in a UK secondary school context. This research involved groups of 2-3 students working around a shared computer to research 'role models', scaffolded by a worksheet that asked open and closed questions regarding individuals, and asked the students to reflect on the quality of the information they were finding. The second study (reported in, Knight, 2016; Knight, Rienties, Littleton, Mitsui, et al., 2017; and, Knight, Rienties, Littleton, Tempelaar, et al., 2017) was conducted in the context of a first year undergraduate class at a Dutch university. This research – the focus of the following sections - involved groups of 2-3 students using a browser add-on, which included a chat functionality, to collaboratively write a report 'for a government advisor' regarding one of two topics (the safety of a herbicide or a food supplement). In the first of these conditions the students were provided with a set of 11 documents on the topic, of varying quality and sub-topic, and in the latter they searched the internet for information. In both contexts the skills required for mature internet use were of interest:

1. *Skill of **integration**, ability to establish connections between prior knowledge and new information, including across documents, and including where claims are inconsistent or contradictory*
2. *Skill of **sourcing**, ability to identify parameters that characterise the author and conditions of production of the information*
3. *Skill of **corroboration**, or the activity of checking accuracy of information against other sources*

(Rouet, 2006, p. 177 emphasis added)

We see the dialogue – through face-to-face interaction (study 1) and the browser chat tool (study 2) – in light of these three skills; how people talk about integrating, sourcing, and corroborating information, while engaged in those actions, provide us with insight into students' 'epistemic stance'. As such, while in both contexts the dialogue was studied in its own right (through transcription of the audio recording, and analysis of the chat logs), other contextual data regarding the browsing and task behaviour was also analysed. In both cases the aim of the analysis undertaken is to understand what is going on: how

students identify particular claims as important or salient and what reasons they give for this; why particular resources might be treated favourably over others; how students come to synthesise information from across sources. Theoretical considerations regarding the role of dialogue, and the context (including resources), thus motivated a particular methodological approach to data collection and analyses.

Analysis in both cases does not focus on attempts to 'code' explicitly for the dimensions of epistemic cognition (certainty, simplicity, source, justification). The intent is not to identify these dimensions in dialogue, and to assess their adaptivity. Instead, the discursive approach entails re-specifying these dimensions in terms of discursive properties that can be identified for example, in a dialogue transcript or chat log. Thus, we see epistemic cognition through the investigation of how students identify particular claims as important or salient over others, requiring a detailed analysis of language regarding "claims" (assertions, topic language, etc.), "sources" (metadata, sourcing language, etc.), and the kinds of justificatory language we might see around these ("because", "I think that…", etc.) as described further below.

In the first study, this approach involved using video and screencast data to investigate the context of utterances made, alongside analysis of the finished written artefact: a worksheet in this case. Analysis of this collection of data provided deeper insight than each individually, or the dialogue alone. For example, one group noted in their worksheet that the BBC was a high quality source; in the screencast data, we observed that they in fact added the term 'BBC' to a query, in order to foreground results from that source. In the second study, the browser add-on logged the chat data. Alongside this data, it also stored the queries each individual made, the pages visited, any text they 'copied' from the pages visited, and their activity on a collaborative document editor. Again, this data has informed our understanding of the dialogue data, and the epistemic nature of the tasks.

3.2.1 Deductive Content Analysis for a Typology of Dialogue

Across both empirical studies discussed here a theorised methodological approach was taken to analyse the dialogue data, drawing on other resources including the task instructions, articles and other media students encountered, and task resources such as the worksheet and browser add-on. In both studies, theory informed the analysis of the data, in developing typologies for understanding how the students approached the tasks. In the first study, the dialogue was analysed both for the presence of 'exploratory' dialogue, and for other relevant epistemic themes (reported in our earlier work). In the second study, a more formalised approach was adopted, that built on deductive content analysis (see, for examples, Elo & Kyngäs, 2008; Hsieh & Shannon, 2005; Mayring, 2000; and Hatch, 2002 for a deductive typology development) to draw out terms for a typology of chat messages.

In this deductive process, theoretic considerations are used to identify key types within the typology and the concepts associated with them. Application of these types can involve the identification of variables or definitions for the typology, and the development of new types where themes emerge from the data that do not fit the a priori characterisations. In this research, three epistemic dialogue types were selected, based on analysis of the research literature and surface analysis of the chat data: topic, source quality, and exploratory chat.

1. The first of these – labelled 'topic' – relates to the topic-content of the material being sought. 'Topic' chat regards the theme (for example, mentioning 'red yeast rice'), the sub-themes (for example, mentioning 'urine'), and other content knowledge around the particular tasks (for example, references to 'health' impacts).

2. A second category 'source quality' relates to use of metadata and source citation, a feature that, for example, Goldman and Scardamalia (2013) identify as important in complex literacy tasks. These terms were selected based on known metadata from the documents and websites encountered (for example, the author name 'Gillam', or the domain name drugs.com), and terms relating to citation ('written by', 'published in', etc.).
3. Finally, the category 'exploratory' includes terms which refer to the kind of meta-discourse Goldman and Scardamalia highlight, and which are related to exploratory dialogue, and epistemic stance ('because', 'I think', 'therefore', etc.).

The intention in any analysis of this kind is not to develop an exhaustive coding scheme as such, but to highlight some structures inherent within the chat data as indicated through the application of a typology of terms. Thus, terms were selected through:

- a priori identification, using prior research and terms directly related to the typology themes, for example the 'exploratory' terms are derived from prior analysis for exploratory terms in the published literature, while in the source quality theme the use of 'written by', 'published in', and known source metadata is derived from prior knowledge of the typology theme;
- a deductive analysis of the materials, including the task instructions, and task-resources, from which further topic and source terms were derived (for example, identification of metadata and topic-terms such as authorship in the documents);
- and finally a deductive analysis of the chat data, to identify further terms, and ensure the chat data aligned with the typology.

This deductive analysis of the chat data indicated an additional category of chat message involving terms around writing activities, including: decisions to copy/paste; concerns around plagiarism; and suggestions to edit the etherpad. Thus a fourth type was included, resulting in four types, Topic, Metadata, Metadiscourse, and Synthesis (ToMMS):

4. Terms identified in messages of this type were labelled 'synthesis', with a theoretically driven hypothesis that those participants who engaged in more discussion that made use of these terms would engage in more coordination of their writing, and that this coordination might be associated with synthesising behaviour, in particular that this discussion might enable participants to integrate the information that they each (separately) found.

In order to conduct this deductive analysis of the chat data, messages containing each term were identified using a concordance-style analysis, and visually inspected. This analysis was conducted by selecting only those messages in which any given term occurred. Such analysis has commonly be conducted using concordance software, which facilitates the exploration of 'Key Words In Context' (KWIC) by displaying words searched for in their original context (typically, showing a sub-portion or whole sentence in which the term is located). This approach is founded in a kind of sociocultural approach that has been deployed in the analysis of dialogue "in action" described both with regard to language as a tool to learn, and (in the context of discursive psychology) a more general tool 'to do' and a lens onto psychological constructs (such as epistemic cognition).

In this way, a sociocultural approach draws on quantitative methods (the counting of the occurrence of particular terms) to inform qualitative analysis (Mercer, 2004). As such, concordance analysis facilitates researchers in testing: "hypotheses about how topics are being carried forward and how meaning is being jointly developed through talk" (Mercer, 2000, p. 69). In the analysis we undertook, terms aligned with the typology described, and as discussed in the following sections, were identified and messages

containing those terms marked. At the preliminary stage messages with these terms occurring in them were visually inspected, in order to explore how the particular term was being used, and its alignment with the conceptual theme (topic, source quality, exploratory). Those terms that were not aligned with the theme were removed. In this way, false-positives are minimised. Not all of the terms identified were widely used, for example 'good page' was used on a single occasion, but broader constructions (for example 'page') were not specific enough in their use, or resulted in results based on quoting the task instructions (for example 'claim' was used in the task instructions, thus instances of 'claim' are often not to do with specific claims, while this is not true of the small number of instances of 'claiming'), thus selection was targeted at identifying those terms best related to the typology themes on both a theoretical basis, and derived from the chat data itself. In addition to the concordance-style analysis, the first 5000 of the 20913 raw messages (i.e., with blank messages, server log messages, etc. intact) were visually inspected for further terms using a top down deductive selection process to select terms associated with the developed themes. Thus, a broader set of terms could be identified than through a priori selection or identification of terms co-occurring with existing target terms alone, thus minimising false-negatives.

3.2.3 Exemplification of concordance approach

In order to confirm that the tasks were giving rise to epistemic commitments the data were manually analysed, with a particular focus on exploration of the chat data for indications of the typology described – of topic, source, exploratory, and (arising from that analysis) synthesis talk. The chat data is particularly foregrounded here as a core feature of the discursive environment in which the tasks were conducted, providing particular insight into the ways in which participants approached the problem presented, and made use of their mediating tools, including their shared dialogue.

Within the analysis sample, Table 3:1 indicates the (non-mutually-exclusive) number of chat messages sent with terms within each element of the typology (with a percentage indicator in brackets). This table shows that messages with one or more exploratory terms in them were most common in both tasks, with topic-terms and source-terms next most frequently occurring in both topics. Note that roughly two thirds of messages contained none of the terms targeted within the typology; some exemplifications of such messages are given below in section 4.2.3.5. As indicated, the typology was developed as an analytic device to capture salient distinctions between particular types of talk of interest to epistemic cognition research, namely: topic content; source qualities; exploratory dialogue; and, arising from the deductive process, synthesis chat pertaining to writing co-ordination.

Table 3:1 – Chat Typology Frequencies in the Food Supplement and Herbicide topic groups

	Food supplement (%)	Herbicide (%)
Exploratory	1276 (20.75)	1046 (18.63)
Source Quality	320 (5.20)	528 (9.40)
Topic	620 (10.08)	243 (4.33)
Synthesis	528 (8.59)	392 (6.98)
No terms	4000 (65.06)	3867 (68.86)

Total messages	6148	5616

Messages including terms of each type were visually inspected, and exemplifications selected. These exemplifications are not intended to provide an exhaustive overview of the kinds of messages sent, nor to imply any kind of 'coding' or classification. Rather they are intended to give the reader insight into the kinds of epistemic chat engaged in, aligned with the typology. The messages are given without preceding or proceeding messages (one exception is noted). While this limits the salient context available for analysis, it provides insight into the level of data captured by key-term identification methods coding at the message level. Moreover, it facilitates the researcher to identify the ways in which terms are used across messages (although their wider structure is obscured) such that comparison between the use of terms in different individual messages may be made. These exemplifications thus illustrate the epistemic nature of the messages including the target-terms in the typology. A set of messages is selected to exemplify the kind of expressions the target-terms appear in; as such, the number of examples provided for each element of the typology, and topic varies. These exemplifications are intended to illustrate the variety of data, and the way in which well-established tools of sociocultural discourse analysis may be applied in a novel context to provide insight on epistemic dialogue.

4.2.3.1 Exploratory chat

Exploratory terms are intended to capture instances of partners explaining their ideas, engaging with the ideas of others, and attempting to build shared understanding. Derived from prior research on exploratory dialogue, and Kärkkäinen's analysis of epistemic stance, instances of 'I think', 'because', 'so', 'maybe', are characteristic of the kinds of language of interest.

Table 3:2 – Typology of Chat Terms

Type	**Terms in typology** (each separated by a bar '|', in some cases a 'stem' is given, e.g. 'conclu' to capture all affixes)
Exploratory[+] – related to sharing and co-positioning with relation to knowledge claims	cuz|because|cause|coz|cos|cus|caus|becuz|I think|I guess|for example|I dont know|I don't know|I guess|I thought|so |if |also |maybe|probably|apparently|aparently|of course|would|must|might|could|will|may|conclu|weigh|agree|therefore

[+]Note: 'best support' is used in quoting the task instructions, and is thus excluded from this analysis; agree is mostly used e.g. "I agree" to partner (rather than, e.g. "these sources agree"); weigh is used as in 'weighing up the arguments'; conclu[de/sion] is used as in "we can conclude that" as is 'therefore' (rather than "the paper has a conclusion", etc.).

For example, in the messages shown in Table 3:3 we see messages exhibiting coordinating (e.g. 1, 4, 9, 12, 18), explaining perspectives (2, 3, 5, 6, 15, 16) and explaining processes or procedures (e.g. 7, 11, 14). Note that only the second of these – explaining perspectives – might constitute the kind of co-construction of common knowledge implicated by exploratory dialogue. The other two types of example might be more accurately described as 'explanatory' in nature, and are perhaps more procedurally focussed. These exemplifications demonstrate the complexities of creating typologies of chat messages on a term or cue-phrase basis. However, they also provide a clear exemplification of the kinds of more exploratory, epistemically salient, dialogue anticipated in such tasks indicating the potential of such analysis in the context of epistemic activities.

Table 3:3 – Example exploratory term messages*

Example	**Message**

1	maybe one can do adventages and the other disadvanatges
2	i think the disadvantages are more important then the advantages
3	because i'm looking at an other website right now that talks about different sideeffects than i mentioned before
4	so how are we going to create a summary together
5	because of the active ingredient which they also use for drugs
6	I think Reuters is more reliable than that site
7	I think its better that we keep looking from different sources since we have more variety
8	yes i think we just just have to write a big conclusion on that rice
9	ok I will just make them in some sort of logical order
10	haha yeah we should but I think that those are the important main point about this red yeast rice
11	maybe we can try to find the best arguments in each article article by article so maybe we should start with reading government urged to act and on
12	if you read the first 5 links and I read the second 5
13	i'm finished and i think the best argument is
14	There are both claims that its harmful and that it is not so we should include both right
15	but I agree it's better than the 2nd to last article that I read
16	So I would say we leave out the first two because they support a different claim than the majority
17	k so what do you think our conclusion should be
18	I just put in everything which I think is important and rewrite it at the end ok

* Typographical errors are retained from the original chat data in these examples and all subsequent examples.

4.2.3.2 Source quality chat

'Source quality' terms related to the use of source metadata and referencing, derived both from known documents (particularly in the Herbicide topic), analysis of webpages (in the Food supplement task), and through analysis of the chat data for citation or referencing related terms. In line with the work on use of source metadata (for example, author, title, etc.) a full analysis of metadiscourse (for example building on analysis by Hyland, 1998, in the context of written texts) is not conducted, as the key interest here is restricted to use of 'evidentials' (in Hyland's framework), i.e. in whether students refer to sources or their qualities directly.

Table 3:4 – Typology of Chat Terms

Type	**Terms in typology** (each separated by a bar '|', in some cases a 'stem' is given, e.g. 'conclu' to capture all affixes)
sourQual# - related to identifying sources, source metadata, and source metadiscourse	written by|prof|publish|author|recent|source| site|good page|bias|scienti|blog|news|cite|sponsor|credible|samsel|senef|indep|reuter |kloor|gillam|duke|powles|gm|fote|foe|friends of the earth|laboratory|bremen|sanders|nutritional sciences|murphy|boobis|fda|method|they sa|\"| webmd| web md| medicinenet| medicine net| lexology| altmedicine| alt medicine| sciencebasedpharmacy| science based pharmacy| drugs.com| medicine.com| wikipedia| emedicinehealth| medicine health| mayo| mercola| maryland| pennstate| medicinenet| medicine net| reuters| ncbi| about.com| NCCAM| ANSES| FDA| efsa| clinicaltrials| dfg.de| consumerlab| umm.edu|webmd |web

	md |medicinenet |medicine net |lexology |altmedicine |alt medicine |sciencebasedpharmacy |science based pharmacy |drugs.com |medicine.com |wikipedia |emedicinehealth |medicine health |mayo |mercola |maryland |pennstate |medicinenet |medicine net |reuters |ncbi |about.com |NCCAM |ANSES |FDA |efsa |clinicaltrials |dfg.de |consumerlab |umm.edu |data|document|article|backed up|evidence|writer|medicine net|study|trust|claiming

[#] Note: The term 'Dr' was rarely used, 'doctor' more so but often in context of the content rather than source; 'by' was used frequently for purposes other than sourcing (e.g. "make headings by..."); similarly 'written' is used variably, and 'written in' (rare) used to refer to the etherpad, 'written by' is used to refer to sources (but is rare); 'old' was not used, nor was outdated or historic; 'new' used to refer e.g. to new tab, etc., while 'recent' was used to refer to sources; 'good' is too widely used (e.g. 'good health' 'good question', 'good job') to use globally; 'good paper', 'good web[site]', etc. not used, although 'good source' is used this is captured by use of 'source' as a keyterm, 'good article' and 'good site' are used rarely but included here; other sentiment terms (e.g. 'great', 'fantastic', 'excellent', 'wonderful', 'brilliant', 'splendid') are either not used, or not used re: sourcing ('great'); 'bad' is used only in the context of topics, not the sources; 'method' is used where participants are evaluating the material in the sources; note the \"" is to catch all instances of quotation. Finally, note url references are not coded here (because they are aggregate as 'link use') although they are related to sourcing discussion.
Data used to refer to "there isn't enough data", etc. 'Claim' is used to refer to source instructions, while 'claiming' is used e.g. "since the people who are tested are the people who are claiming that it's bad".

Analysis of source-terms in chat messages (as in Table 3:5) indicated that sources were often referred to in reference to the specific claims within the source, but often using a generic term such as "a site" or "an article" rather than more specific metadata (e.g. 1, 18, 19), or referencing sources in passing (e.g. 9). Also evidenced are references to source and authorial qualities, for example particular publication venues (e.g. 4,5,14,20,27), or authorial characteristics e.g. 'scientists', 24,25,26,28) or both (e.g. 11, and the combination of 21-22, which were sequential messages sent by one dyad). Finally, generic comments on source quality (e.g. 3,8,12) suggesting looking for 'good sources' or citing a range of sources, are seen. Note that lines 21 and 22 were sequential, and by different partners in a pair, indicating a small exemplar of source-based discussion. Again here we see, along with some less-relevant messages, some very clear exemplifications of the kinds of chat messages around source qualities and source meta-data of interest to this research. Throughout these messages coordinated activity around sources and sourcing can be seen, with specific and explicit evaluation of source qualities (generally or specifically) given in various places.

Table 3:5 – Example source quality term messages

Example	Message
1	this is what a site said
2	just have to copy everything from this site
3	remember to reference to reliable sources
4	is wikipedia enough reliable
5	I think Reuters is more reliable than that site
6	does the source where u got the animal study from say how high the doses were
7	all the sites that i found references the same site
8	I think its better that we keep looking from different sources since we have more variety
9	so u keep looknig reuters
10	I also find an article that is against the use of rice should i include it too or only articles that support the rice safety
11	lets back it up with scientific research rather than sum reuters news
12	lets collect the sources at the bottom and then were done i guess

13	its just a drugs company but they say they screen their yeast for impurities
14	kinda commercial site though it seems
15	but I'm missin something more scientific
16	shall we spilt the articles
17	after reading each article should we make a summary of it and then combine everything together
18	oh damn the next article is telling us that the first one isnt true hahaha this sucks
19	all 4 articles that i have read are pro claim
20	Yeah there's one of a scientific magazine which is not biased
21	I'm not sure whether the critisism from the science media center is scientific enough the way it is phrased just sound really colloquial
22	Yeah but it are all scientists
23	in the 3rd article they went to say that the average was 44 but the Uk scientists denounced the study as unrealiable Us biotech giant monsanto says that 40 years of independent assessments says it chemical doesn't pose a risk to human health
24	oh and some of the sourceses are crap one is written by a blogger we are only suppose to take those who seem accurate you know what i mean
25	but like all professors are claiming that the sample is not proper due to the fact that the real data was not published
26	what are Friends of the Earth and GM Freeze NGO s
27	I skipped one because it was a blogg
28	Article 10 is basically about a discover blogger named Keith Kloor who says that the research of one Carey Gillam is bullshit because she has done the reseach in a field which is not her expert field because she doesn't have any diplomas on that field or whatsoever

4.2.3.3 Topic

'Topic' terms related to the content of the material, and its topics and sub-topics and the specific claims contained within documents (including statistical claims). Analysis of messages including a topic-term indicated that topic-terms appeared in a range of contexts indicated in Table 3:7.

Table 3:6 – Typology of Chat Terms

Type	**Terms in typology** (each separated by a bar '|', in some cases a 'stem' is given, e.g. 'conclu' to capture all affixes)
Topic* - related to content terms, specific claims, and key-words related to the content theme	statin|ryr|red yeast|rice|citrin|health|heart|diseas|amount|monacolin|contamin|cholesterol|drug|liver|lovastatin|medicin|french|kidney|china|muscl|supplemenet|ingredi|toxi|prescri|diet|ferment glyphosate|roundup|monsato|urine|salt|2 microgram|2.5 litres|residues|farmer|weed|environ|cytochrome|obes|xenobiotic|gastro|heart|diab|depres|autis|infert|cancer|alzh|entropy|west|inflam|agri|health|disease|stati|herbicid|volunteers|44 per cent|44%|44 percent|18 countries|samples|10%|90%|plant|crop|weedkil|pesticide

* Note: These topic terms are taken from visual inspection of a tfidf output from the Food supplement and Herbicide topic chat corpora, and (particularly in the Herbicide case) prior knowledge of the topic materials.

As can be seen in the exemplifications below, some topic-terms were used in a rather general context (for example, 4-7, 10, 13, 20), discussing the broad problem and theme of the task. At other times, specific points are being made, drawing on general information rather than specific sources or claims (for example, 1-3, 9, 11), while in other cases specific claims are drawn out and noted or/and discussed (for example, 8, 12, 14-19, 21). These messages exemplify a more topic-focussed content, with a more or less general target in each message. It is interesting that, perhaps due to the length of the messages, these topic-focussed terms tend not to co-occur with sourcing, exploratory, or synthesis terms. That is, for example, within individual messages there is little evidence of explanation (exploratory terms) or weighing up specific or contested claims in light of their role in the written output (synthesis terms) or their provenance (source quality terms). This absence is perhaps due to the focus on individual messages rather than longer extracts, however as the following section will illustrate, the exemplifications provided here do in fact give a reasonable representation of the kinds of chat that occurred.

Table 3:7 – Example topic term messages

Example	Message
1	it is dangerous as works simmilar to drugs affecting liver
2	so it may cause harm if somebody has liver problems etc
3	we can add sometrhing about contamination
4	do we like the red yeast rice
5	it is not the worst but definitely exposes one to health risks
6	did you ever hear of that rice D
7	i guess we can do something like first looking at the potential health risks
8	Animal studies have been conducted in China using high doses of red yeast rice products No damage to the kidneys liver or other organs were demonstrated in these studies
9	What do you already have I have that it can be dangerous if it contains monacolin K and the side effects of it
10	should we talk about the confusion drug or supplement
11	Yep it's some kind of toxic stuff that fucks up your kidney
12	wait this one That's not long o Glyphosates Suppression of Cytochrome P450 Enzymes and Amino Acid Biosynthesis by the Gut Microbiome Pathways to Modern Diseases
13	So do your articles state that the herbicide is damaging or harmless
14	The first two articles imply almost everyone in Europe has traces of glyphosate
15	well i kinda got it from disease severity but that might be wrong since its more about plant
16	The fifth article implies farmers are more susceptible
17	In the environmental benefits part it says 'environmentally benign''
18	it seems pretty unlikely it leads to ALS Alzheimer's MS cancer and autism
19	Glyphosate is less acutely toxic than common chemicals such as sodium chloride or aspirin with an LD50 for rats greater than 5 g kg1
20	it is used as weed killer
21	theres one document about the health risks to humans

4.2.3.4 Synthesis

The 'synthesis' type was derived from an analysis of the chat messages, which foregrounded that a category of salient chat around the coordination of writing and synthesis was not accounted for by the three initial types.

Table 3:8 – Typology of Chat Terms

Type	**Terms in typology** (each separated by a bar '|', in some cases a 'stem' is given, e.g. 'conclu' to capture all affixes)
Synthesis – related to coordination regarding writing[α]	structure|phrase|formulate|plag|conclus|intro|heading|write|own words|summary|task pad|copy|paste

[α] Note 'plag' is to capture plagiarism, plagiarised, etc. (and misspellings) – this has been manually checked for false positives with satisfactory results.

Analysis of messages including synthesis-terms or text-coordinating language (shown in Table 3:9) indicated references to the report structure – often noting introductions, or conclusions specifically – (for example, 1,3,12,18,19), with attempts to coordinate writing (9,10,11) the taking and sharing of notes (2,4,5,13,14,16,17,) or/and use of paraphrasing, copy-pasting, and citation management (6,7,8,15). Messages within this facet of the typology exemplify the coordination of activity, attempts to understand the task-context and requirements, and to address how well those requirements were being met. As such, these messages are epistemically salient, considering the kinds of information needed to present, how this information should be presented and contextualised towards the task goals, and how to co-ordinate knowledge sharing within the pairs – these considerations thus broadly relate to the simplicity/complexity of knowledge, and the role of the self/other in sourcing knowledge.

Table 3:9 – Example synthesis term messages

Message	**Example**
1	how should we structure the report
2	I'm just going to write some random notes down then we can structure it once we have enough info
3	so we have the basic intro
4	we need to rewrite everything
5	lets not look for more information but rather rewrite vererything and the come to a conslusion
6	so we can copy paste the links to say we got the info from there
7	Do you think we need to 'summarise' it into our own words or what is the meaning of all this that we just know how to look things up
8	we are going to prison for plagarism
9	okay ill write it down
10	how much do we have to write
11	yeah do you want me to write something about it being illegal in USA
12	first introduction what it is then prositiv and negativ arguments and afterwards we judge it
13	after reading each article should we make a summary of it and then combine everything together
14	yeah its way more reliable because its a real experiment Dont read what they did try to find conclusions or at least thats what im doing
15	yeah but i dont think we should add any of our own words just phrase what is said by other

16	so should we start with just writing in the task pad what each article is about so we both know
17	do you want to copy some stuff into the taskapad
18	we could just make a short introduction stating what it is
19	What should the conclusion be

4.2.3.5 Messages without target-terms

In order to investigate the nature of the messages without any target keyterms in them, a sample of 10% of such messages (Food supplement n = 400, Herbicide n = 386) was visually inspected. Messages in this subset often had a number of (non-mutually-exclusive) features including that they:

- Were very short (e.g. “yes”, “good”, “okay”), or omitted elements which would have been identified in the typology (e.g. “but this is short term”)
- Were not in English (frequently German or Dutch)[1]
- Were off-topic (e.g. “im the coolest one int he [sic] room”)
- Included terms with a high incidence of false-positives (e.g. “okay you're right then just jump right into the arguments” is excluded because task instructions were frequently copied into the text-chat in their entirety, and included the term)
- Included typographical errors (e.g. “i thinjk “)
- Included alternative terms to those included in the typology (e.g. instead of toxic, “there is this poisonous thing in there sometimes”)
- Involved some task discussion (e.g. “this project is boring”, “i have no idea what the hell we are doing here”, “press submit”)

Given the deductive process of term selection inevitably some less frequently used terms of interest have not been identified, or their inclusion would increase the numbers of false-positives. In some cases, this means that epistemically interesting messages are not identified (e.g. “let us collect a lot of facts and then in the end look for the strongest ones“).

4.2.3.6 Section Summary

As highlighted in Table 3:1, across the full set of messages a substantial number (roughly two thirds) contained none of the terms contained within the typology types. The discussion above indicates some of the kinds of messages sent within this subset of messages, demonstrating frequent messages with less salience for the analytic focus in this work, alongside some less clear cases. Across the exemplifications shown in sections 4.2.3.1-4.2.3.4 there are clear indicators of epistemic dialogue, aligned with the typology described. These cases demonstrate the potential value of such term-based approaches in identifying the types of ways in which participants engage in dialogue around information seeking and processing tasks. However, across these cases we also see a depth of nuance that is not captured through simple summary counts of typology-aligned-term instances. Nonetheless such counts provide additional insight, allowing us to engage in both an analysis of raw chat messages, and quantitative methods (particularly counts) which may be taken to inform our understanding of the qualitative analysis (see, for example, Mercer, 2004). In addition, this close analysis of the messages indicates a number of what one might consider ‘false positives’ (in the exemplifications), and ‘false negatives’ (described briefly in section 4.2.3.5). As described in the earlier sections, the intent of the typology development is not to capture all possible types of dialogue, or the breadth of nuance, but

1 While the students were taught in English, and typically work in English, some communicated via the chat program in their mother tongue. Frequently these messages also included English terms (particularly around the topic, or source features).

rather to develop an analytic device to distinguish between particular terms in the use of dialogue. Thus, additional types may emerge from the data (for example, as indicated above, off task talk) with less salience to the analytic approach. Finally, while the analysis provided here indicates clear epistemic dialogue, with differences between the ways in which terms are used apparent in the text, they offer coarser insight into the interactional nature of the dialogue – the ways in which the participants shared and built their knowledge together. To address this, the following section will engage in a closer analysis of a subset of groups, discussing their interaction with the tools (indicated through analysis of trace data), and each other (indicated through the chat data).

3.2.4 Analysis of chat excerpts

In addition to concordance style analysis of chat data, a closer analysis was conducted on a small subset of projects' trace data both as a component of the deductive process, and, aligned with sociocultural approaches to discourse, as a means to draw on both fine and coarser grain analysis of the dialogue data in understanding the interactional processes. This analysis was conducted in order to examine the epistemic features of the task behaviour as foregrounded through the chat data.

An approach that has been developed to investigate the temporal development of dialogue as, and in, context is to look for how terms are repeated over a dialogue episode, and from external resources (such as webpages) into a dialogue, to understand how "speakers can jointly, co-operatively create cohesion in…their speech" (Mercer, 2000, p. 62). Analysis of terms in the typology aids this investigation by providing information on how those terms are used within individual messages, drawing on that wider context. This approach is grounded in a 'systemic functional linguistics' perspective, holding that texts reveal contextual features through their genre (See Halliday, Hasan and Christie 1989), that in the case of dialogue "is created anew in every interaction between a speaker and listener or writer and reader. From this perspective, we must take account of listeners and readers as well as speakers and writers, who create meanings together" (Mercer, 2000, p. 21). In the empirical work to which our approach is applied key elements of 'context' include the dialogue itself, the webpages that participants visited and from which they drew, and the final written outputs (worksheet answers in study 1, and written report in study 2).

Analysis of this kind allows the researcher to investigate both broad similarities (for example, overall levels of chat), and differences (for example, the presence of particular types within the typology) across groups. The closer analysis of the data also reveals more nuanced differences in the ways in which the participants interact with the tools, each other, and the task. These examples highlight the epistemic nature of the tasks, and how this is seen in the chat data; they also foreground the complexity of analysis of such data as a form of behavioural indicator associated with learning outcomes. Across the three excerpts there are clear differences, observed in the kinds of language drawn on, specifically drawn out through the use of this more extended analysis of dialogue above the single utterance/message level.

3.2.4.1 Excerpt 1

In the Herbicide task, this group wrote a particularly high quality report, evaluating and synthesising the sources to cover the full range of sub-topics. The group wrote a relatively long text, although analysis of the authorship indicates that one partner appears to have contributed most of the content (perhaps due to one participant pasting larger chunks of text into the pad).

The group began the task segment expressing uncertainty regarding the task:

userID	projectID	localTime	Content

610	325	17:39:50	what to we actually have to do
611	325	17:40:04	i have no clue haha
610	325	17:40:13	thats unfortunate
610	325	17:40:15	Meneithe
610	325	17:40:17	R
611	325	17:40:20	read the articles

After about 8 minutes exploring the task pad, and each looking at documents and the instructions, one suggests:

userID	projectID	localTime	Content
611	325	17:47:20	do you want to do the first 5 and i do the last 5

They then engage in a stretch of reading time. Towards the end of the task the pair then discuss how best to represent and summarise the information found, including how to organise 'pro' and 'con arguments and whether to synthesise information or primarily use quotation:

userID	projectID	localTime	Content
610	325	18:01:06	do you want to sort yout stuff to my list of pros and cons
611	325	18:02:42	did you summarize them
611	325	18:02:49	or copy passages
610	325	18:03:00	copy passages
610	325	18:03:13	no way gonna summarize that takes way to long
610	325	18:03:19	or are we supposed to do that
611	325	18:04:16	i ahve no clue

Then they discuss how to weigh up the articles, and collate them into groups to give their opinion on. In this end stage they also note an inter-textual tie between two articles, saying: "well one of the pros references to one of the con articles". We thus see some attempts to synthesise information (identifying ties between articles, and grouping claims), with a concern to evaluate or give opinions on the claims given.

userID	projectID	localTime	content
610	325	18:05:33	we have to write why we think there the best too
610	325	18:05:37	damn
611	325	18:05:53	the arguments or what
610	325	18:05:58	yeah i think so
610	325	18:06:08	but we cant do that for all of them
611	325	18:06:17	maybe as a conclusion
610	325	18:07:02	check the task pad
610	325	18:07:07	i put the description
610	325	18:07:28	we have to decide which ones are the best arguments
611	325	18:07:36	yeah but still for every single argument
611	325	18:07:38	thats too much work
611	325	18:07:42	oh alright

611	325	18:08:33	well one of the pros references to one of the con articles
610	325	18:08:50	we have to grouop them
611	325	18:10:01	ehm for which aspects
610	325	18:11:34	lets stat putting the arguments into groups and wirte our opinions

3.2.4.2 Excerpt 2

Also in the Herbicide task, this group wrote a relatively less sophisticated report, mostly listing claims from the sources, with little integration or evaluation of them. This group mostly drew on a single article, referring to it in the written output, chat, or copying directly from the pages.

The pair start off reading the articles, both reading the first two articles independently. They then discuss dividing the articles between them, briefly noting that the first article (a scientific report) is better than the first (a press release):

userID	projectID	localTime	Content
692	369	10:11:31	what do you think of these articles so far
693	369	10:12:08	do we need to read all of them
692	369	10:12:54	I think so but we can divide the articles between each other
693	369	10:13:15	Nice
692	369	10:13:41	which articles have you read so far I've the first 2
692	369	10:13:51	Read
693	369	10:13:54	mee too
693	369	10:14:32	so we can divide other articles
692	369	10:14:34	second one is better than the first one would you agree
692	369	10:14:56	Yh
693	369	10:14:59	Yes
693	369	10:15:21	argree1
693	369	10:15:31	Agree
692	369	10:16:19	ok so how shall we divide it I'll read the bottom 4 and you read the rest
693	369	10:17:28	Okbut i don't know what will we do after finish the reading
693	369	10:17:40	give a summary
692	369	10:17:49	of each one
693	369	10:18:04	maybe the best one
692	369	10:18:40	I think because we only have until 1045

Each then flags a 'best' article from the articles they have read:

userID	projectID	localTime	content
692	369	10:25:17	ok I've read the bottom 4 and the best one out of them all is the 2nd to last article
693	369	10:25:39	I read the fifith article it looks nice

They then discuss how to deal with the length given the short amount of time, deciding to:

userID	projectID	localTime	content
692	369	10:38:35	I think we should go for article 9 and copy and paste the main claim from the article

We thus see a focus on a single article here, rather than on trying to find (synthesise and evaluate) claims from multiple different sources. This suggests that the single source is taken as the authoritative resource from which to draw material, yet there is little discussion of the qualities of that source (in terms of metadata or content), and the ways in which the source relates or does not relate to other sources (and their metadata and content).

3.2.4.3 Excerpt 3

This group, in the Food supplement task, also wrote a relatively less sophisticated report. Their log-data indicated that they chatted or copied text from five webpages (from the 17 that they viewed), and although they made a number of queries between them (30 google searches), these contained only 14 unique terms in total, indicating relatively little vocabulary richness.

The group starts by suggesting that they should create a short summary of the nature of red yeast rice:

userID	projectID	localTime	Content
1100	586	10:11:50	maybe we should get a quick summary together of what red yeast rice is
1099	586	10:12:22	alright

Going on to suggest they might summarise in the task pad (or 'main board'), using quotations to do so:

userID	projectID	localTime	Content
1100	586	10:18:25	so it also has positive effects
1099	586	10:18:52	lets just summ them up on the main board
1100	586	10:19:32	ok
1100	586	10:19:55	u know how to quote stuff
1100	586	10:20:05	so then we can just copz paste it
1099	586	10:21:46	lets do just like this

They send some messages about what red yeast rice is, concluding with the message:

userID	projectID	localTime	Content
1100	586	10:39:21	its kiond of a medicine with some side effects

3.2.4.4 Excerpt 4

This group, also in the Food supplement task, again wrote a relatively less sophisticated report. In contrast to excerpt 4, the pair viewed a small number of pages (n = 9), only making use (through the chat, task pad, or copying from) 2 pages, which were shared early on in the chat data. In contrast while they issued a similar number of queries to other groups (n = 27), there were more unique terms in these (n = 31) indicating a wider vocabulary use across the queries issued.

The group started by both finding and sharing a useful article. They briefly discus which article to draw from but make a selection without any discussion or explicit evaluation of the source quality or its content:

userID	projectID	localTime	Content
1187	633	12:47:22	i found a good article from the university of maryland medical center
1187	633	12:48:15	http//ummedu/health/medical/altmed/supplement/redyeastrice

1187	633	12:49:03	read and we can start to write
1188	633	12:49:10	http//wwwivlproductscom/HealthLibrary/HealthConcerns/HeartHealth/RedYeastRiceforCholesterolTheProsandCons/
1188	633	12:49:18	i found this
1187	633	12:50:07	ok i will read
1188	633	12:50:30	can we just copy and paste D
1188	633	12:50:47	Hahaha
1187	633	12:50:48	from you site or my
1188	633	12:50:55	as you want
1188	633	12:51:00	yours is better i think
1187	633	12:51:18	okay let's make a good copy oopen the task pad
1188	633	12:51:36	we can choose the most intersting topics and just past them

Finally, they divide the content of the site they are drawing on, negotiating that they will each summarise some of the points made ("I do the first three points and you the other two"):

Then they divide the points they want to make, so one does 3 the other 2

userID	projectID	localTime	Content
1187	633	13:06:37	because it's to long on te website
1188	633	13:07:06	ok let's do a summary
1187	633	13:07:25	maybe just put the name
1188	633	13:07:31	i do the first three points and you the other two

4. Discussion and Conclusions

When students engage in learning they must engage with the claims of others, coming to evaluate and reconcile them to take some stance. To do this, they must understand the discursive context of those claims, the purpose for which they are made, and the norms and practices that surround their limits and strengths. Typical accounts of epistemic cognition have focussed on the developmental or cognitive processes through which students come to learn these practices. However, a range of work, including our own, has acknowledged the discursive nature of epistemic cognition. In this paper we have drawn upon this prior empirical and theoretical work to develop a novel approach to the analysis of discourse data in epistemic tasks. This analysis is characterised in relation to: (1) The content or topicality of claims made in a dialogue; (2) the ways that sources are referred to and used to make these claims; (3) the expression and development of epistemic stance through the use of metadiscourse or exploratory dialogue; (4) the ways this stance is developed across sources, towards some task-oriented ends.

We exemplify this approach drawing particularly on data from a task in which participants engaged in a computer-based information-seeking task. These exemplifications illustrate both the nature of the particular datasets, and the application of the approach to that data, showing how concordance-style analysis, alongside analysis of longer excerpts, can be drawn on to understand epistemic features in the discourse. As in other sociocultural discourse analysis (Mercer, 2004), the intent of the analytic device is to provide insight into language as a tool for *doing.* This analysis supports both close reading of the data, and the use of quantitative approaches to provide further insight into how the language is being used; in this case, based on quantification of the presence of typology terms, and in other analyses allowing

comparison of this quantification across groups of students. The approach is not intended to provide a coding scheme or to capture all features of a text, but rather to provide the analyst with an approach to developing a heuristic for understanding salient *discursive* epistemic features. While extant work in epistemic cognition has drawn on dialogue data, no overarching approach or method for the treatment of that data has been described. This paper's core contribution is thus to provide an analytic approach to the epistemic features of dialogue. In so doing, we exemplify an approach to taking seriously the discursive properties of epistemic cognition, motivated by a theorised social account of the construct.

Acknowledgements

This work was undertaken while the first author was undertaking his PhD at the Open University, UK, and first presented in that thesis (Knight, 2016). Thanks are extended to colleagues at the Open University, particularly Professor Bart Rienties, and to collaborators Dr Dirk Tempelaar, Dr Chirag Shah, and Matthew Mitsui. Thanks are also extended to Dr Chi-Hé Elder for helpful discussions regarding earlier versions of this paper.

References

Edwards, D. (2005). Discursive Psychology. In K. L. Fitch & R. E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 257–273). Erlbaum.

Edwards, D., & Mercer, N. (1987). *Common knowledge: the development of understanding in the classroom*. London, UK: Routledge.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Goldman, S. R., & Scardamalia, M. (2013). Managing, Understanding, Applying, and Creating Knowledge in the Information Age: Next-Generation Challenges and Opportunities. *Cognition and Instruction*, *31*(2), 255–269. https://doi.org/10.1080/10824669.2013.773217

Greeno, J., Collins, A., & Resnick, L. B. (1992). Cognition and learning. In B. Berliner & R. Calfee (Eds.), *Handbook of Educational Psychology*. New York. Retrieved from http://www.wou.edu/~girodm/611/GCR.pdf

Halliday, M. A. K., Hasan, R., & Christie, F. (1989). *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Oxford University Press, USA.

Hammer, D., & Elby, A. (2003). Tapping Epistemological Resources for Learning Physics. *Journal of the Learning Sciences*, *12*(1), 53–90. https://doi.org/10.1207/S15327809JLS1201_3

Hatch, J. A. (2002). Chapter 4: Analyzing Qualitative Data - Typographical Analysis. In *Doing Qualitative Research in Education Settings* (pp. 152–161). SUNY Press.

Hornik, K., Mair, P., Johannes, R., Geiger, W., Buchta, C., & Feinerer, I. (2013). The textcat Package for n-Gram Based Text Categorization in R. *Journal of Statistical Software*, *52*(6), 1–17. Retrieved from http://www.jstatsoft.org/v52/i06/

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hutchison, P., & Hammer, D. (2010). Attending to student epistemological framing in a science classroom. *Science Education*, *94*(3), 506–524. https://doi.org/10.1002/sce.20373

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*(4), 437–455. https://doi.org/10.1016/S0378-2166(98)00009-5

Kärkkäinen, E. (2006). Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk*, *26*(6), 699–731. https://doi.org/10.1515/TEXT.2006.029

Knight, S. (2016). *Developing Learning Analytics for Epistemic Commitments in a Collaborative Information Seeking Environment* (PhD). Open University, Milton Keynes, UK. Retrieved from http://oro.open.ac.uk/46143/

Knight, S., Allen, L., Littleton, K., Rienties, B., & Tempelaar, D. T. (2016). Writing Analytics for Epistemic Features of Student Writing. Presented at the International Conference of the Learning Sciences, Singapore.

Knight, S., Arastoopour, G., Williamson Shaffer, D., Buckingham Shum, S., & Littleton, K. (2014). Epistemic Networks for Epistemic Commitments. In *International Conference of the Learning Sciences*. Boulder, CO: International Society of the Learning Sciences. Retrieved from http://oro.open.ac.uk/39254/

Knight, S., & Littleton, K. (2017). Socialising Epistemic Cognition. *Educational Research Review*, *21*, 17–32. https://doi.org/10.1016/j.edurev.2017.02.003

Knight, S., & Mercer, N. (2015). The role of exploratory talk in classroom search engine tasks. *Technology, Pedagogy and Education*, *24*(3), 303–319. https://doi.org/10.1 08 0/14 75 939 X .2014.931884

Knight, S., & Mercer, N. (2016). The role of collaborative, epistemic discourse in classroom information seeking tasks. *Technology, Pedagogy and Education*, *26*(1), 33–50. https://doi.org/10.1080/1475939X.2016.1159978

Knight, S., Rienties, B., Littleton, K., Mitsui, M., Tempelaar, D. T., & Shah, C. (2017). The relationship of (perceived) epistemic cognition to interaction with resources on the internet. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2017.04.014

Knight, S., Rienties, B., Littleton, K., Tempelaar, D. T., Mitsui, M., & Shah, C. (2017). The Orchestration of a Collaborative Information Seeking Learning Task. *Information Retrieval*. https://doi.org/10.1007/s10791-017-9304-z

Littleton, K., & Mercer, N. (2013). *Interthinking: putting talk to work*. London: Routledge.

Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: evaluating and learning online information. *Metacognition and Learning*, *5*(1), 67–90. https://doi.org/10.1007/s11409-009-9048-2

Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1089

Mercer, N. (2000). *Words & Minds: How we use language to think together*. Oxon: Routledge.

Mercer, N. (2004). Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, *1*(2), 137–168. https://doi.org/10.1558/japl.v1i2.137

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), 359–377. https://doi.org/10.1080/0141192040100016896 89

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, *1*(1), 12–21. https://doi.org/10.1016/j.lcsi.2012.03.001

Mercer, N., & Littleton, K. (2007). *Dialogue and the Development of Children's Thinking: A Sociocultural Approach* (New edition). Oxon: Routledge.

Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education*, *20*(6), 507–528. https://doi.org/10.2167/le678.0

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's Talk and the Development of Reasoning in the Classroom. *British Educational Research Journal*, *25*(1), 95–111. https://doi.org/10.1080/0141192990250107

Österholm, M. (2009). Theories of epistemological beliefs and communication: A unifying attempt. In *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 275–264). Retrieved from http://umu.diva-portal.org/smash/record.jsf?pid=diva2:228322

Potter, J., & Edwards, D. (1999). Social Representations and Discursive Psychology: From Cognition to Action. *Culture & Psychology*, *5*(4), 447–458. https://doi.org/10.1177/1354067X9954004

Reznitskaya, A., & Gregory, M. (2013). Student Thought and Classroom Language: Examining the Mechanisms of Change in Dialogic Teaching. *Educational Psychologist*, *48*(2), 114–133. https://doi.org/10.1080/00461520.2013.775898

Rojas-Drummond, S., Littleton, K., Hernández, F., & Zúñiga, M. (2010). Dialogical interactions among peers in collaborative writing contexts. In K. Littleton & C. Howe (Eds.), *Educational dialogues: Understanding and promoting productive interaction* (pp. 128–148). Abingdon, Oxon: Routledge.

Rosenberg, S., Hammer, D., & Phelan, J. (2006). Multiple Epistemological Coherences in an Eighth-Grade Discussion of the Rock Cycle. *Journal of the Learning Sciences*, *15*(2), 261–292. https://doi.org/10.1207/s15327809jls1502_4

Rouet, J.-F. (2006). *The Skills of Document Use: From Text Comprehension to Web-Based Learning* (First Edition edition). Mahwah, NJ: Routledge.

van Dijk, T. A. (2014). *Discourse and Knowledge: A Sociocognitive Approach*. Cambridge University Press.

Wegerif, R. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning*, *1*, 143–157. https://doi.org/10.1007/s11412-006-6840-8