Teacher questions, wait time, and student output in classroom interaction in EMI science classes: An interdisciplinary view

Jiangshan An ✉
Purdue University Fort Wayne, USA
https://orcid.org/0000-0003-4214-4283
anj@pfw.edu

Ann Childs
University of Oxford, UK
https://orcid.org/0000-0001-5918-739X
ann.childs@education.ox.ac.uk

Abstract

Past research has often shown a lack of student output in English medium instruction (EMI) classes (e.g., An et al., 2021; Lo & Macaro, 2012) and this study seeks to identify possible reasons. Guided by literature on wait time (Rowe, 1986) and teacher higher-order thinking questions (Chin, 2007), this study explores whether these two pedagogical moves have the same impact on classroom interaction in EMI science classes. 30 EMI science lessons were recorded from seven EMI high school programs in China, taught by 15 native speakers of English to homogenous groups of Chinese students. Correlation tests showed that when there was more wait time after a teacher question, the students produced lengthier responses with more linguistic complexity, took up more talk time, and asked more questions. However, greater use of teacher higher-order thinking questions, coded by Chin's (2007) framework of constructivist questions, did not correlate with any student output measures. This suggests that wait time may be a more effective factor leading to more student output in EMI classes than asking higher-order thinking questions. Qualitative

analysis showed teachers' follow-up moves may have also played a role in the limited success of higher-order thinking questions.

1. Introduction

In recent years, English medium instruction (EMI) programs have been rapidly growing across the world from higher education to secondary and primary education (An & Murphy, 2018; Macaro et al., 2018). These programs adopt English to teach subject matter in contexts where the local population typically do not speak English as their first language (L1, Macaro et al., 2018). In Europe, they are usually referred to as content and language integrated learning (CLIL) and elsewhere as EMI.

Research in science education and language education has established that interaction is an important mechanism for learning to take place (Long, 1996; Mortimer & Scott, 2003). Although studies in EMI have often described the classroom interaction in such classrooms, few have analyzed the impact of specific pedagogical moves on student participation. This study aims to fill this gap by exploring how pedagogical moves such as the use of higher-order thinking questions and wait time influence student output in EMI science classes in foreign high school programs in China.

2. Literature review

2.1. The role of interaction for learning

The significance of interaction in learning can find its roots in sociocultural theory (SCT). As Vygotsky (1986) states, cognitive development originates from social contexts and proceeds to individual mental activity. During social interaction, a learner can be assisted by a more competent other to accomplish a task which is beyond the learner's current ability. This process is termed *scaffolding* (Wood et al., 1976). This conceptualization of learning means in classrooms interaction is an important channel for learning to take place. The socio-constructivist view of learning (Erdogan & Campbell, 2008), consistent with SCT, further highlights that students should be given ample opportunities to articulate their thinking (Mercer, 2004). In second language acquisition (SLA), it is now well accepted that language development needs not only input but also output where learners can test their hypotheses of language forms and notice the gap between their

interlanguage and the target forms (Swain, 1985). Long's (1996) interaction hypothesis argued that the modified input and feedback that occur during negotiation of meaning are particularly beneficial for second or foreign language (L2) development, highlighting again the significance of interaction.

2.2. Teacher questions

Teacher questioning is a key tool to shift classroom discourse to be more interactive. In a science classroom featuring constructivist teaching approaches, teacher questions often aim to encourage students to elaborate on their ideas, discuss various points of view and thus promote higher-order thinking (Chin, 2007). Such questions can elicit more substantial student responses in full sentences, benefiting science learning (Chin, 2006, 2007; van Zee & Minstrell, 1997). Constructivist teaching is often contrasted with teaching by transmission where teacher questions often elicit only restricted student responses consisting of pre-determined "single detached words" (Chin, 2006, p. 1317) which typically only require lower-order thinking (van Zee & Minstrell, 1997).

2.3. Wait time

The wait time a teacher leaves after asking a question and before a student response is a component of teacher questioning strategy that could also impact student responses (Black & Wiliam, 1998). Rowe's (1974) influential work identified two types of wait time. Wait time I is the period of time which immediately follows a teacher's question but before a student answers and Wait time II is the time period following a student's answer before the teacher responds. In this study we are focusing only on Wait Time I because there was little evidence of Wait time II in our data. Rowe's work found that teachers normally leave an average of less than one second of wait time after asking a question (Rowe, 1974). Studies later found that an increased wait time, to a threshold of three seconds or more, gives students more time to think about the questions and is associated with positive changes in the classroom interaction patterns, including increased number and length of student utterances (Swift & Gooding, 1983) and student answers being "supported by evidence and logical argument" (Rowe, 1986, p. 44). Tobin (1987) further argued that average wait time greater than three seconds led to higher achievement in learning.

2.4. Classroom interaction and teacher questioning in EMI classes

In the EMI literature, studies often find limited classroom interaction (e.g., An et al., 2021; Lo & Macaro, 2012). Teacher questioning behavior may be one reason.

What is commonly found is a pattern of mostly recall questions and rare use of higher-order thinking questions (Sopia et al., 2010; Yip et al., 2007). As one of the few studies that compared the types of teacher questions and the student output elicited, Llinares and Pascual Peña (2015) found in CLIL history classes in Madrid that 65.84% of teacher questions were recall questions, with questions for eliciting facts producing the simplest and shortest responses. In addition, questions asking for reasons and metacognitive questions generated the most complex responses. In contrast, Dalton-Puffer (2007) found in CLIL lessons in Austria that while questions for facts were predominant at 89%, short student responses featuring single noun phrases persisted independent of the type of questions. This, as Dalton-Puffer speculated, could be because students "need more time to think and formulate" (p. 117), signaling a need of more wait time. Thus, evidence remains inconclusive as to whether higher-order thinking questions elicit more substantial and complicated student responses in EMI classes, as claimed for L1 classes. In addition, there is little research on wait time in EMI contexts. Given the dual challenges of learning subject knowledge and the L2, one could speculate that wait time is more necessary in EMI classes to allow longer student utterances with more complexity.

2.5. Teachers in EMI classes

In EMI studies, EMI teachers' own English proficiency has often been called into question and identified as a reason for the prevalent use of closed and lower-order thinking questions (Sopia et al., 2010; Yip et al., 2007). Thus, one could ask whether EMI teachers with a high level of English proficiency would use questions differently and thus elicit more student responses. While acknowledging that the term *native speaker teacher* (NST) is problematic (see An et al., 2021, for a detailed discussion), we decided to retain it to refer to the teachers in our study as it is the teachers' high English proficiency that allows for the exploration of the relationship between teacher questions and student output without the restriction of the teachers' own English language proficiency. The research questions are as follows:

1. What are the patterns of teacher higher-order thinking questions, wait time, and student output in the classroom interaction in EMI science classes taught by NSTs in foreign high school programs in China?
2. What are the relationships between teacher higher-order thinking questions, wait time and student output in these classrooms?

3. Methods

3.1. Research context

This study was situated in EMI foreign high school programs in China. These programs often adopt an Anglophone high school curriculum and have foreign teachers instruct local Chinese students through English only. The students are typically aged 16-18 years old, and usually plan to study overseas in English-speaking countries for their tertiary education.

3.2. Sample

The data of this study came from seven EMI foreign high school programs across China, featuring 15 NSTs and 308 Chinese students. Convenience sampling was adopted due to accessibility issues and only the schools that gave access were recruited. The authors did not have a personal relationship with the participants. Consistent efforts were made to ensure a reasonable representation of the target school programs, including geographical location and the type of curriculum taught, as shown in Table 1 below.

Table 1 Teacher background

Province	School	Curriculum	T	Subject	Gender	Age	Nationality
	Sch 1	Canadian British Columbia	T1	Chemistry	F	33	Canadian
			T2	Physics	M	54	Canadian
			T3	Biology	F	52	Canadian
	Sch 2	UK IGCSE, AS, A2	T4	Biology	M	29	American
Province A	Sch 3	Canadian British Columbia	T5	Physics	M	25	Canadian
			T6	Chemistry	M	59	Canadian
			T7	Biology	F	24	Canadian
	Sch 4	Canadian Alberta	T8	Physics	M	56	Canadian
Province B	Sch 5	American AP	T9	Biology	M	34	American
	Sch 6	IB	T10	Biology	M	36	American
			T11	Physics	M	24	Canadian
Province C	Sch 7	Canadian British Columbia	T12	Chemistry	F	23	Canadian
			T13	Biology	F	31	Canadian
			T14	Biology	F	29	Canadian
		UK IGCSE, AS, A2	T15	Biology	M	32	British

As shown in the teacher background questionnaire, all 15 teachers held at least a bachelor's degree and were certified teachers in their home countries. All of them identified English as their most proficient language, thus confirming their NST status, and stated not having a functioning proficiency of Mandarin.

The teachers commented in interviews that most of the students had strong science knowledge and an intermediate level of English proficiency. Given

a lack of standard exams in these programs, students' answers to three items in a student questionnaire were used to understand how students' English proficiency might impact on the output they produce in class, as shown in Table 2. A 5-point Likert scale was used, including choices of 1 – *strongly disagree*, 2 – *disagree*, 3 – *neutral*, 4 – *agree*, and 5 – *strongly agree*.

Given the normal distribution of the answers from all three questions in the 15 classes (Kolmogorov-Smirnov statistic greater than .05) and the assumption of homogeneity of variances met (Levene's test's static greater than .05), ANOVA was run. Results showed no significant differences among the 15 classes for all three questions. This may indicate that any differences in student output was not due to the variation in the students' English proficiency in different classes.

Table 2 Students' self-reported impact of English on classroom interaction

	Student questionnaire items – how English proficiency impacts interaction		
	1) I very often don't understand the teacher in science classes.	2) In science classes, sometimes I know the answers to teachers' questions, but I don't answer because I am afraid of speaking in English.	3) In science classes, sometimes I know the answers to teachers' questions, but I don't answer because I don't know how to phrase it in English.
M (SD)	2.13 (0.85)	2.45 (1.04)	2.79 (1.08)
ANOVA	$F_{(13, 197)} = 1.66$ $p > .05$	$F_{(13, 196)} = 1.19$ $p > .05$	$F_{(13, 197)} = 2.12$ $p > .05$

3.3. Data collection

Video recordings of two consecutive lessons for each teacher were conducted by the first author. A naturalistic non-intervention observation approach was adopted. The 30 lessons observed covered a wide range of topics and each lesson lasted between 45 minutes to one hour. A later screening of the lessons excluded two lessons, including T9's second lesson where a lengthy student debate activity took place and T10's second lesson consisting of one teacher monologue followed by group discussion. Before the observations, information sheets were given to the participants, and they were debriefed on the purpose and use of the data. All the lessons recorded were from classes where consent was obtained.

3.4. Data analysis

The video recordings of the lessons were entered into NVivo 11 software, where teacher-whole class interaction in each lesson was transcribed verbatim.

3.4.1. Quantitative analysis

The quantitative analysis aimed to identify the overall pattern of wait time, teacher higher-order thinking questions, student output and correlations among the three constructs.

Wait time was defined as pauses of any length after a teacher's question and before a student's response during teacher-whole class interaction. All wait time was coded in NVivo to the 0.00 seconds and the software produced the total length of wait time in each lesson. Teacher questions were also coded, which produced the number of teacher questions in each lesson. The average length of wait time per teacher question in each lesson was used to represent the degree that wait time was used in each lesson.

Teacher questions were further coded using Chin's (2007) framework of constructivist teacher questioning approaches. Questions that match these types were considered higher-order thinking questions. Chin's (2007) framework can be found in Table 3.

Table 3 Chin's (2007) framework of constructivist questioning approaches

Type of constructivist questioning approaches	Functions	Sub-type constructivist questioning strategies
1. *Socratic questioning*	Elicit students' reasoning based on prior knowledge rather than directly transmitting knowledge to them.	*Pumping* – the teacher asks for more information from students to foster students' talk rather than giving the answer directly. *Reflective toss* – the teacher throws back the responsibility of providing feedback to a student's response to the same or a different student. *Constructive challenge* – when students provide an incorrect answer, the teacher responds with a question to lead students to realize their own misconceptions.
2. *Verbal jigsaw*	Consolidate students' linguistic knowledge of science terminology to form declarative statements	*Association of key words and phrases* serves to elicit key scientific vocabulary from students for the formulation of declarative knowledge and build up a mental framework, especially when there is a high number of technical terms involved. *Verbal cloze* – the teacher leaves out blanks in their sentences for students to fill in.
3. *Semantic tapestry*	Help students connect ideas together and construct cohesive understandings	*Multi-pronged questioning* – the teacher asks students to approach one issue from different angles, for example, through processing and producing information in textual descriptions and in drawings. *Stimulating multimodal thinking* – the teacher asks students to switch between a variety of modes of thinking, for example, through visual images, linguistics or symbolic resources or formulas, to solve a problem. *Framing & zooming* – the teacher adjusts the questions depending on the kind of thinking to be elicited, e.g., at the macro /observational level or micro/molecular level.
4. *Framing*	Use questions to frame a problem to structure the discussion.	*Question-based prelude* – an expository preface to help students see the structure of the information introduced subsequently. *Question-based outlines* – the teacher provides a set of outline sub-questions to break down an overarching question into smaller steps. *Question-based summary* – a summary in a brief question-and-answer format to reinforce the key concepts.

Chin's framework allowed a fine-grained analysis of a wide range of higher-order thinking questions specific to science classes to advance students' thinking through dialogue. The percentage of higher-order questions to the number of teacher questions in each lesson was calculated to represent the degree to which higher-order thinking questions were used. The number of sub-types of questions was also identified to describe the varieties of higher-order thinking questions. Recall questions were also coded.

To measure student output in each lesson, four parameters were used: 1) the average turn length of student responses after a teacher question; 2) the noun verb ratio in student responses to teacher questions; 3) the number of student questions asked; 4) the time percentage of student talk to total teacher-whole class interaction time. Parameters 1) and 4) were adopted from Lo and Macaro's (2012) study on classroom interaction in EMI secondary schools in Hong Kong. Parameter 1) reflects the degree the students provide substantial elaborations. Parameter 2) was adopted from Macaro et al.'s (2016) work and represents the complexity level of the linguistic structure of student responses. In science classes, more verbs indicate more complete descriptions of science processes as they typically involve verbs. Parameter 3) represents the degree students initiate dialogue, a particular type of student output. These four measures were obtained through coding student talk in the lessons in NVivo.

To ensure the coding was accurate, 10% of the lessons (i.e., three lessons) were randomly selected to be coded again on all measures by another researcher. This resulted in an inter-rater reliability of .78, indicating a reasonable level of reliability (Robson, 2002).

To answer Research Question 2 (RQ2), correlation tests were run in SPSS to determine correlations between the use of wait time, teacher higher-order thinking questions, and the four measures of student output in each lesson.

3.4.2. Qualitative analysis

In answering RQ2, qualitative analysis was also conducted through examining the lesson transcripts to understand how the correlation results manifested themselves in the classrooms (Borkowska, 2011).

In understanding how teacher questions impacted student output, the use of follow-up questions was also analyzed, particularly when the initial questions did not elicit full responses. In addition to Chin's (2007) framework, Tang's (2021) framework of five types of follow-up moves in science classes was also consulted. These moves include *extend*, *probe*, *paraphrase*, *reflective toss*, and *constructive challenge*. *Extend* refers to teachers' follow-up question to push students to move forward their reasoning until a full explanation is given to account

for a phenomenon. *Probe* refers to moves that push students' reasoning backwards from an outcome to the cause. The moves *reflective toss* and *constructivist challenge* are also identified in Chin's (2007) framework.

4. Results

4.1. RQ1: Patterns of teacher higher order thinking questions, wait time and student output

The descriptive statistics of all the measures in the 28 lessons are shown in Table 4.

Table 4 Patterns of teacher questions, wait time and student output

Constructs	Variables/measurements	*M*	*SD*
Teacher higher order thinking question	Percentage of higher order thinking questions to all teacher questions (%)	46.83	24.10
Wait time	Average length of wait time after a teacher question (in secs)	1.01	1.23
Student output	Average turn length of student responses to a teacher question (in secs)	3.30	1.58
	Noun verb ratio in student responses	5.19: 1	3.45
	Number of student questions	2.46	3.28
	Time percentage of student talk to teacher-whole class interaction (%)	10.06	7.55

4.1.1. Teacher question types

As background information, on average 54.13 questions were asked by the teachers in a lesson and one teacher question occurred every 49.63 seconds during teacher-whole class interaction time. This shows first that the NSTs asked questions frequently. Almost half, 46.83%, were higher-order thinking questions by Chin's (2007) definition. However, only limited types of higher-order thinking questions were used. The breakdown of each type is shown in Table 5. *Pumping* was the most widely used type, accounting for 24.61% of all teacher questions, which is 52.55% of all higher-order thinking questions. Other types were rather rare. As shown in Table 6, the use of recall questions was low, 9.75%.

Table 5 Percentages of different types of higher-order thinking questions

Question types	Total number of occurrences	Percentage to the total number of teacher questions
Higher-order thinking questions	742	46.83%
a. Socratic questioning	425	27.45%
• pumping	381	24.61%
• constructive challenge	10	0.65%
• reflective toss	0	0.00%
b. Verbal jigsaw	221	14.28%
• association of key words and phrases	220	14.21%
• verbal cloze	1	0.06%
c. Semantic tapestry	85	5.49%
• multi-pronged questioning	29	1.87%
• stimulating multi-model thinking	44	2.84%
• framing and zooming	12	0.78%
d. Framing	11	0.71%
• question-based prelude	0	0.00%
• question-based outline	11	0.71%
• question-based summary	0	0.00%

Table 6 Percentage of recall questions

Questions considered as lower-order thinking	Total number of occurrences	Percentage to the total number of teacher questions
Recall questions	151	9.75%

4.1.2. Wait time

Wait time after teacher questions had a rather short average length of 1.01 seconds per lesson, showing the teachers generally did not leave long wait times. However, there was a wide range of average wait time across the lessons, as shown by the standard deviation of 1.23 seconds, indicating some degree of variation in the teachers' practices.

4.1.3. Student output

The average turn length of student responses to teacher questions was rather short, 3.30 seconds. This indicates that the students generally did not provide substantial output answering teacher questions. The noun verb ratio in student responses, 5.19:1, showed a strong noun-oriented nature, indicating limited use of verbs. Student questions were overall rare and occurred 2.46 times on average per lesson. The time percentage of student talk averaged 10.06% of teacher whole-class interaction time, showing overall limited student participation.

4.2. RQ2: Correlations between wait time, teacher question types and student output

Based on the scatterplots generated in SPSS, linearity and homogeneity of variance were met for the bivariate correlation model to be used. Based on the Kolmogorov-Smirnov statistic, all variables had non-normal distribution except two – the percentage of teacher higher-order thinking questions to all questions and the average turn length of student responses, as shown in Table 7.

Table 7 Normality of the measures

	Wait time	Percentage of teacher higher-order thinking questions	Turn length of student responses	Noun verb ratio in student responses	Number of student questions	Time percentage of student talk
Kolmogorov-Smirnov statistic Sig	0.00	0.20	0.07	0.03	0.00	0.03
Normal distribution	No	Yes	Yes	No	No	No

Results of correlations are shown in Table 8 and Table 9. Spearman's Rho were run except for the correlation between teacher higher-order thinking questions and turn length of student responses, where Pearson was used.

Table 8 Correlations between wait time, teacher questions and measures of student output

	Student output			
	Turn length of student responses	Noun verb ratio in student responses	Number of student questions	Time percentage of student talk
Wait time	$r = .46^*, p < .05$	$r = -.45^*, p < .05$	$r = .42^*, p < .05$	$r = .43^*, p < .05$
Teacher higher-order thinking questions	$r = -.07, p > .05$ (Pearson)	$r = .16, p > .05$	$r = .18, p > .05$	$r = -.10, p > .05$

Table 9 Correlation between wait time and teacher questions

	Wait time
Teacher higher-order thinking questions	$r = .30, p > .05$

The results show that wait time has a significant moderate positive correlation with all four measures of student output while teachers' higher-order thinking questions did not have a significant correlation with any student output measures. The absence of correlation between teacher higher-order thinking questions and

wait time shows that when the teachers asked questions that posed a higher cognitive demand, they did not leave more wait time.

4.3. RQ2: Qualitative results of how teacher questions and wait time impacted student output

Complementing the quantitative results, the qualitative analysis provided insights into finer details of how wait time and teacher higher-order thinking questions were used and impacted student output.

4.3.1. Excerpts 1 & 2: Use of extended wait time to elicit more student output

While wait time was generally short, when there was more substantial wait time, the students tended to produce more substantial answers to both higher-order thinking and lower-order thinking questions. Excerpt 1 from T7's biology lesson on plant structure demonstrates how extended wait time, after a higher-order thinking question, was followed by an extensive student response:

Lesson excerpt 1

Turn	Timespan	Content	Speaker
37	14:10-14:22	So, the stems grow upwards, and they branch outwards to maximize the total surface area of the leaves. So why would a plant want to grow upwards?	T
39	14:22-14:40	[wait time]	
40	14:40-14:42	Kira?	T
41	14:42-14:50	Err, err, the more upwards, there is less shadows, so the plant can get more energy from the sun.	S
42	14:50-15:00	Good. So, the more upwards it grows, the higher it gets, the more access to light it can have, the less shadows.	T

In introducing "stem," the teacher asked a *pumping* question in Turn 37: *So why would a plant want to grow upwards*? to ask students to speculate rather than giving students the information directly, placing a relatively higher cognitive demand on them. Then there was a lengthy wait time of eight seconds, which was followed by a rather substantial response from a student with a turn length of eight seconds in full sentences with both agents and verbs (e.g., *is, can get*). Following the student's answer, in Turn 42 the teacher provided a paraphrase of the students' answer in the target language forms. It could be argued that the substantial student output in this excerpt was a result of both an open-ended *pumping* question that aims to foster students' talk and the generous use of wait time.

Excerpt 2 from T13's biology lesson on continental drift theory demonstrates how extended wait time after a recall question also led to substantial student output with sophisticated linguistic structure:

Lesson excerpt 2

Turn	Timespan	Content	Speaker
62	17:11-17:25	They are made out of, they are made out of plate tectonics. Plate tectonics. Right? Now how does a volcano or an earthquake occur on earth? What happens?	T
63	17:26-17:29	[wait time]	
64	17:30-17:31	Ivy?	T
65	17:32-17:43	Volcano and um earthquake happen at the age (edge) of the continents which um help to separates (separate) um the continents from each other.	S
66	17:43-17:46	Help to separate? What do you mean by separate?	T
67	17:46-17:49	[wait time]	
68	17:49-17:59	Um because um the, especially the lava from the volcano uh came out and uh it forms new rock. Then uh the.	S
69	18:00-18:05	But how does that happen? So, um what happens to the plates? They what?	T
70	18:05-18:07	Move.	Ss
71	18:07-18:11	Yeah. They move. They get in contact with each other. Right? ...	T

This exchange took place at the beginning of this class where the teacher was revising previous content. In Turn 62, the teacher asked a recall question about how a volcano or an earthquake occurs in revising tectonic plate theory. Although recall questions typically require a lower level of cognitive demand, the teacher still provided three seconds of wait time in Turn 63, which might be because this question asked for a complete description of a cause of a phenomenon. In Turn 65, a student was able to give an initial response of 11 seconds in a full sentence with both agents and verbs (e.g., *happen, help, separate*). However, this answer was not a fully correct answer. Then the teacher asked a follow-up question in Turn 66: *Help to separate? What do you mean by separate?*, which focuses on the part that needed further thought. This question was followed by another extended wait time of three seconds, given in Turn 67. In Turn 68, the student provided another lengthy response of 10 seconds, again in full sentences using the verbs *came out* and *forms*. However, this answer described the outcome of volcano eruption rather than the cause. In Turn 69, the teacher continued the dialogue with another follow-up question to push for the exact cause: *but how does that happen?* and *what happens to the plates?*. This seemed to be a *probing* follow-up move (Tang, 2021) as it pushes students to identify the underlying cause for a phenomenon. This elicited the key word *move* from the

students in reference to the cause. The teacher then provided feedback confirming the cause being the movement of tectonic plates leading to collision between them. Here, the generous use of wait time at different points of this exchange with a chain of follow-up questions appeared to have allowed students the time needed to recall relevant information and organize substantial answers in the L2.

4.3.2. Excerpts 3 & 4: Challenges of higher-order thinking questions to elicit student output

As the quantitative results show, the use of more higher-order thinking teacher questions did not elicit more substantial student output. Examination of the lesson excerpts shows often initial higher-order thinking questions received incomplete student answers, and there was a lack of follow-up questions or effective follow-up questions by the teacher to push students to elaborate their answers. This pattern also coincides with the lack of variety of higher-order thinking questions identified in the quantitative results in that the follow-up questions did not seem to make full use of the different higher-order thinking question strategies. Excerpt 3 is an example from T15's biology lesson on genetic modification:

Lesson excerpt 3

Turn	Timespan	Content	Speaker
34	19:03-19:06	Debbie, do you think identical twins have the same fingerprints?	T
35	19:06-19:07	No.	S
36	19:07-19:08	Why?	T
37	19:08-19:11	Er, because it's just no.	S
38	19:11-19:27	Just no. Well yeah, alright, they don't. Fingerprints are not genetic. You don't get your thumbprints or your fingerprints from your genes. Fingerprints actually arrive when you're growing inside the womb and it's just your skin folding randomly.	T

In Turn 34, the teacher asked a *pumping* question to elicit students' ideas about whether identical twins have the same fingerprints. After the student's short answer *No* in Turn 35, the teacher asked a follow-up *pumping* question: *Why* to invite elaboration from the student. This follow-up question is also a *probing* move (Tang, 2021) as it aimed to elicit the underlying principle of an outcome. However, this probing move did not elicit an elaboration from the student, as shown in Turn 37. Possible reasons could be that the student was experiencing language difficulties and was only able to essentially repeat the same answer: *it's just no.* It could be that she did not know the key word that was needed, *genetic*, or did not know how to organize her answer with an appropriate sentence structure, such as *XX is not genetic* or *XX is not decided by genes*. After this short

student turn, the teacher in Turn 38 immediately provided a full explanation himself: *fingerprints are not genetic.* Here it could be argued that another follow-up *probing* question that focuses on eliciting the key word, *genes* or *genetic*, could be helpful, for example, *what decides fingerprints?* The teacher may also model the use of key language items as part of the follow-up question. An example is *identical twins have the same eye color because eye color is decided by genes. So why do you think identical twins do not have the same fingerprints?* The first part serves as a modelling of a possible sentence structure: *A is decided by B* as well as the key word *genes*. This may scaffold students' use of language to provide a full answer.

Excerpt 4 below is from T11's first physics lesson on sound waves. In the previous lessons, the concepts and diagrams of sound waves were introduced, as shown in Figures 1 and 2, and the teacher conducted an experiment with an open-open tube with two turning forks, one of 512 Hertz and one of 256 Hertz to demonstrate resonance. In this lesson, the teacher briefly repeated this experiment, where the tuning fork of 512 Hertz had resonance, and asked the students if an open-closed tube was used to achieve resonance whether a longer tube or shorter tube would be needed:

Lesson excerpt 4

| Turn	Timespan	Content	Speaker
11	13:11-14:17	. . . This is 512 Hertz. It works with the open-open tube. Now it's a closed tube, but it doesn't work. Do I need a longer tube or a shorter tube than this?	T
12	14:17-14:20	[wait time]	
13	14:20-14:23	What is your answer? Josephine?	T
14	14:23-14:24	Shorter maybe.	S
15	14:25-14:27	Maybe shorter, OK. Are you imagining the wave, the wave here? [T pointing to the image of 'm=1' in Figure 1]	T
16	14:27-14:28	Yes.	S
17	14:28-15:28	Yes, you are imagining the wave here, OK, do you remember what the other one looks like? So, the previous one, you can fit in half a wavelength. Half wavelength for this L [T pointing to Figure 1]. OK? Here [T showing Figure 2], we can fit in a quarter of the wavelength.	T
18	15:28-15:29	Er, longer!	S
19	15:29-18:57	OK, you are changing your mind now? Does it need longer? Alright we will test it out here. [T conducted an experiment by pouring water into a graduated cylinder, which served as an open-ended tube, to change the length of the tube.]	T
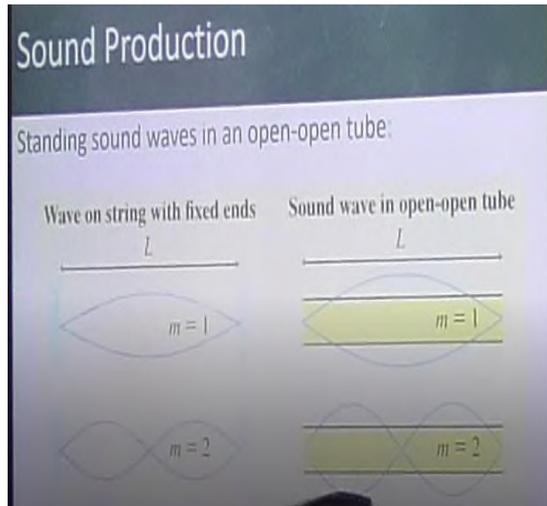
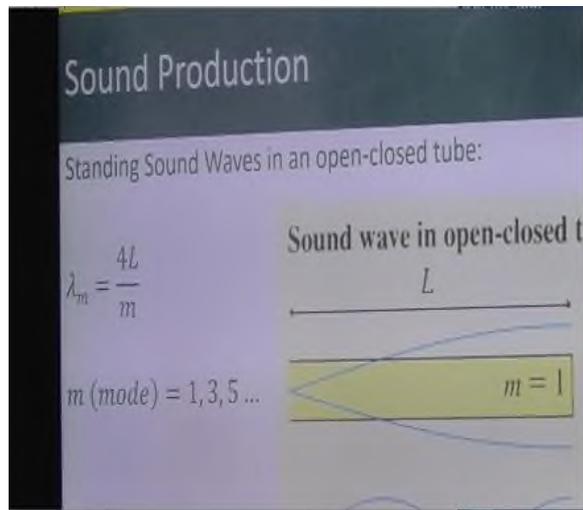**Figure 1** Standing sound waves in an open-open tube

Figure 2 Standing sound waves in an open-closed tube

In Excerpt 4, the teacher asked a *pumping* question in Turn 11 to invite students to make a hypothesis of a new scenario, that is, to obtain resonance with the same tuning fork whether the open-closed tube should be longer or shorter than the open-open tube. Then, a lengthy wait time of three seconds was given in Turn 12. This led to the student's one-word answer, *shorter*. Then the teacher repeated back *maybe shorter* but does not indicate if the student's answer is correct. He then asked a yes/no follow-up question, which referred to Figure 1 to confirm with the student the reason for her answer. His elaboration of

the difference between Figure 1 and Figure 2 in Turn 17 seemed to be an effort to lead students to think more and possibly point to a contradiction. This led the student to quickly change her answer to *longer*. The teacher then conducted an experiment, which proved the student's first answer, *shorter*, was correct. In this exchange, a number of follow-up questions might have been helpful in eliciting students' reasoning behind the one-word answers, *shorter* and *longer*. First, a *why probing* follow-up move (Tang, 2021) might have been useful to elicit the student's elaboration of the principle behind her answer. It might also be possible that the student did not know the answer and guessed *shorter*. Then, with no feedback from the teacher about whether her answer was correct, compounded by subsequent questions from the teacher, the student changed it to *longer*. This indicates that it might be helpful here for the teacher to give feedback, what Chin (2006) calls "accepting" the student's answer and then use questions to elicit the reasoning behind it. If the student does not genuinely know the answer, the teacher could use a *reflective toss* to elicit other students' ideas, for example, whether they agree or not and ask them to elaborate further, involving more students in a richer discussion. Finally, the teacher also could have invited the student to explain her answers by using sound wave diagrams for open-open tubes and open-end tubes with verbal explanations, thus forming a *multi-pronged* questioning episode where the student uses different modalities. This case demonstrates possible missed opportunities of follow-up questioning to address the initial higher-order thinking question.

5. Discussion

This study explored the patterns and relationships of teacher higher-order thinking questions, wait time, and student output in EMI science classes taught by NSTs in the foreign high school programs in China to understand the pedagogical factors impacting classroom interaction in EMI classes.

5.1. RQ1: Patterns of teacher higher-order thinking questions, wait time, and student output

The finding that half of the teacher questions were higher-order thinking questions, with recall questions occupying only a small proportion, clearly contrasts with previous findings featuring low use of higher-order thinking questions and a dominance of recall questions, where the teachers' low English proficiency was often considered a factor. This shows that when EMI teachers possess high English proficiency, they might be more confident in opening up conversations to collectively construct knowledge with students on complicated subject matter and adopt a constructivist teaching approach with a more dialogic nature.

Despite the use of more higher-order thinking questions, however, there was a limited variety. This suggests a possible lack of repertoire of discursive strategies from the teachers to guide students' thinking through dialogues. One reason could be that *pumping*, the most commonly used type, by Chin's (2007) definition, is a more straightforward form of constructivist questions. The minimal use of *constructive challenge* shows when a student gave incorrect answers, the teacher rarely asked him/her or other students to re-think their incorrect answers and led them to work out the answers on their own. The absence of *reflective toss* means that when a student provided a response, the teachers never asked students to evaluate or comment on this response, thus redirecting such responsibility back to the students. As Chin (2007) discussed, each of the questioning approaches possesses a special and meaningful function in contributing to constructivist teaching. Thus, this lack of variety means possible missed learning opportunities for students to realize their misconceptions and discuss a range of views.

The limited overall use of wait time is similar to what was typically found in L1 science classrooms (Rowe, 1974). While we acknowledge that the use of wait time depends on many factors, some of which are cultural (OECD, 2005), the consistent findings across different contexts seem to show that leaving more substantial wait time may be a challenge for most teachers. This is true even in EMI classes where wait time may be more needed for students to think about questions and phrase answers about subject knowledge in an L2.

The patterns of student output reflect a limited degree of student participation. The short average student turn length and the high noun-verb ratio suggest a prevalent use of short noun-oriented answers and limited degree of articulation of science processes, where verbs would be typically required. The rare incidents of student questions show students seldom initiated interaction. Together with the overall low average time percentage of student talk, unfavorable conditions for science learning and language learning were revealed (Chin, 2007; Long, 1996).

5.2. RQ2: Relationships between higher-order thinking questions, wait time and student output

One of the foremost findings of this study is that wait time seemed to be a stronger factor leading to student output of more quantity and quality whereas the use of higher-order thinking questions did not necessarily achieve the same effect. The moderate positive correlation between wait time and all four student output measures suggests that when the teachers did give more wait time, the students were able to produce lengthier output with more complicated linguistic structures involving verbs instead of single-noun answers, talk more and initiate more questions themselves. This effect was regardless of the type of the

questions, as Excerpts 1 and 2 demonstrate. Thus, this finding adds to the existing literature in the L1 classrooms (Rowe, 1974; Tobin, 1987) that in EMI classes more extended wait time can also lead to positive changes to classroom interaction. The moderate level of correlation perhaps indicates a heightened need for wait time for students to think and phrase answers due to the dual cognitive challenges in EMI science classes (An & Thomas, 2021). Given the limited capacity of our working memory (Sweller, 1998), students' working memory may well be overloaded in EMI classes. Thus, wait time is perhaps critically important in EMI classes to allow more substantial responses. This study also shows that more use of wait time in EMI classes may create an atmosphere which signals that the teacher values students' ideas, thus encouraging student questions. This was also observed in L1 classrooms (Samiroden, 1983).

While wait time was shown to be beneficial, the lack of correlation between it and higher-order thinking questions indicates that the teachers did not seem to coordinate the use of wait time with the types of questions they asked. This means the students were not given sufficient time needed to answer higher-order thinking questions. Due to the more complex thinking processes requested, higher-order thinking questions may also place a higher demand on language use. The students may need to create their own language in explaining their reasoning, as compared to likely recycling or reciting the language they received in answering recall questions. Thus, from a language perspective, the lack of longer wait time after higher-order thinking questions may also have inhibited the students from producing substantial answers. Literature on wait time for lower-order questions shows that, although some authors (e.g., Tobin, 1987) question the need for longer wait times for these questions, others (e.g., Ingram & Elliott, 2016) suggest that, even for low level questions, more wait time may also be needed. The findings of this study reinforce the argument that wait time leads to longer and more complex student responses regardless of the question types. This could be because in EMI classes wait time after lower-order thinking questions may be helpful if the language barrier causes challenges to students' responses, as demonstrated by Excerpt 3.

While previous literature argued higher-order thinking questions tend to elicit more substantial and complex student responses (Chin, 2006; Llinares & Pascual Peña, 2015), it was not the case in this study. Apart from the limited use of wait time and the lack of variety of the higher-order thinking questions, another reason could be a lack of effective follow-up questions. The issues of variety and follow-up questions, however, are intertwined. As Excerpts 3 and 4 show, initial higher-order thinking questions, typically *pumping* questions, often did not receive a full answer, and there were often missed opportunities for other varieties of higher-order thinking questions to be asked as follow-up questions.

In Excerpt 4, the single-word answers *shorte*r and *longer* are not sufficient to demonstrate a good understanding and, as described in the results section, various questioning approaches might have been useful to lead students to provide more explanations.

In using follow-up moves to scaffold extended dialogues, this study shows that in EMI science classes such moves need to scaffold both the development of science ideas and the use of appropriate language to describe these ideas. While follow-up questions have been well established in L1 science classes as helpful for pushing students to elaborate on their thinking (Mortimer & Scott, 2003; Tang, 2021), the follow-up moves discussed are typically centered around the science content. However, in EMI contexts language may well inhibit students' ability to elaborate. Thus, multiple follow-up questions may be needed to help students build both science understanding and language. As demonstrated in Excerpt 3, a single *why* follow-up question may not be sufficient in eliciting a further response, particularly when the student struggles to use the appropriate linguistic structure to describe their reasoning. In this case, the teacher may ask more follow-up questions to elicit key words or model the use of key language items before asking the student to give a full answer, examples of which are given in Excerpt 3. This incorporation of the language aspect is another key implication of this study, where we argue that follow-up moves scaffolding the language constitute an additional dimension that needs to be addressed in EMI classes. As shown in this study, higher-order thinking questions do not necessarily elicit student output of more quantity and quality. Thus, follow-up moves that model or elicit key language items are particularly needed. However, given the intertwined nature of language and content, teachers also need to be cautious about modelling the target language without answering the question themselves, which would defeat the purpose of constructivist questioning.

6. Conclusion

This study showed that EMI teachers' high English proficiency may lead to more higher-order thinking questions, and extended wait time may be an effective pedagogical move to elicit lengthier and more complex student responses. However, higher-order thinking questions may not always elicit the kinds of student output that is expected, and a wide range of questioning approaches and multiple effective follow-up questions may be necessary in building extensive dialogues. More research is needed in various contexts to identify effective pedagogical moves enabling more classroom interaction in EMI classes, thus helping achieve the dual goals of EMI.

References

An, J., Macaro, E., & Childs, A. (2021). Classroom interaction in EMI high schools: Do teachers who are native speakers of English make a difference ? *System*, *98,* 102482. https://doi.org/10.1016/j.system.2021.102482

An, J., & Murphy, V. (2018). *English as a medium of instruction in primary schools in South America: A review of the evidence.* A report commissioned by the Oxford University Press.

An, J., & Thomas, N. (2021). Students' beliefs about the role of interaction for science learning and language learning in EMI science classes: Evidence from high schools in China. *Linguistics and Education*, *65*, 100972. https://doi.org/10.1016/j.linged.2021.100972

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. King's College London School of Education.

Borkowska, K. (2011). Approaches to studying classroom discourse: Introduction. In S. Walsh (Ed.), *Exploring classroom discourse language in action* (pp. 67-89). Routledge.

Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education*, *28*(11), 1315-1346.

Chin, C. (2007). Teacher questioning in science classrooms: approaches that stimulate productive thinking. *Journal of Research in Science Teaching*, *44*(6), 815-843.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. John Benjamins.

Erdogan, I., & Campbell, T. (2008). Teacher questioning and interaction patterns in classrooms facilitated with differing levels of constructivist teaching practices. *International Journal of Science Education*, *30*(14), 1891-1914.

Ingram, J., & Elliott, V. (2016). A critical analysis of the role of wait time in classroom interactions and the effects on student and teacher interactional behaviors. *Cambridge Journal of Education*, *46*(1), 37-53.

Llinares, A., & Pascual Peña, I. (2015). A genre approach to the effect of academic questions on CLIL students' language production. *Language and Education*, *29*(1), 15-30.

Lo, Y. Y., & Macaro, E. (2012). The medium of instruction and classroom interaction: Evidence from Hong Kong secondary schools. *International Journal of Bilingual Education and Bilingualism*, *15*(1), 29-52.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, *51*(1), 36-76.

Macaro, E., Graham, S., & Woore, R. (2016). *Improving foreign language teaching: Towards a research-based curriculum and pedagogy*. Routledge.

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

Mortimer, E. F., & Scott, P. (2003). *Meaning making in secondary science classrooms*. Open University Press.

OECD. (2005). Formative assessment: Improving learning in secondary classrooms. *Assessment*, *29*(November), 282. https://www.oecd.org/dataoecd/19/31/35661078.pdf

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed.). Blackwell.

Rowe, M. B. (1974). Pausing phenomena: Influence on the quality of instruction. *Journal of Psycholinguistic Research*, *3*(3), 203-224.

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, *37*(1), 43-50.

Samiroden, W. D. (1983). *The effects of higher cognitive level questions wait time ranges by biology student teachers on student achievement and perception of teacher effectiveness.* [Unpublished doctoral dissertation, Oregon State University].

Sopia, Y., Ong, T., Hashimah, A., Sadiah, B., & Lai, Y. Y. (2010). Teaching science through English: Engaging pupils cognitively. *International CLIL Research Journal*, *1*(3), 46-59.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.

Sweller, J. (1998). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.

Swift, J. N., & Gooding, C. (1983). Interaction of wait time feedback and questioning instruction on middle school science teaching. *Journal of Research in Science Teaching*, *20*(8), 721-730.

Tang, K. S. (2021). *Discourse strategies for science teaching & learning: Research and practice*. Routledge.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research*, *57*(1), 69-95.

van Zee, E. H., & Minstrell, J. (1997). Reflective discourse: Developing shared understandings in a physics classroom. *International Journal of Science Education*, *19*, 209-228.

Vygotsky, L. S. (1986). *The collected work of L. S. Vygotsky. Volume 1: Thinking and speaking*. Plenum.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89-100

Yip, D. Y., Coyle, D., & Tsang, W. (2007). Evaluation of the effects of the medium of instruction on science learning of Hong Kong secondary students: Instructional activities in science lessons. *Education Journal*, *35*(2), 78-107.