The Medium is the Message: How Lecture Q&A Channel Structure Shapes Student Questioning

Meenakshi Mittal*
University of California, Berkeley
meenakshi.mittal@berkeley.edu

Kaden Tang*
University of California, Berkeley
kaden.tang1@berkeley.edu

Christopher Mach*
University of California, Berkeley
mach_chris@berkeley.edu

Narges Norouzi
University of California, Berkeley
norouzi@berkeley.edu

ABSTRACT

Student questions provide a vital signal of how learning unfolds, yet the nature of these questions varies systematically across delivery media. We analyze student help-seeking behavior across two contrasting lecture environments: a synchronous, public Q&A channel where students compete for limited "fixed-time" airwaves, and an asynchronous, private AI lecture support tool that supports "flexible-time" engagement. We introduce an automated, human-validated labeling scheme to characterize over 4,500 student questions across three dimensions: intent type, reasoning explicitness, and confidence. Our application of this scheme suggests that while synchronous channels seem to reinforce a fixed-time/variable-learning bottleneck, the asynchronous medium is more consistent with a flexible-time/fixed-learning model. When inquiry is decoupled from the live lecture stream, the asynchronous channel is associated with denser, more conversational question streams (1.93 turns vs. 1.06 on average) and a significant increase in definition-oriented question types. We show that medium structure appears to meaningfully shape the length and volume of student questioning, shifting question-asking from survival in a live queue toward more iterative conceptual engagement.

Keywords

Question-Answering, Lecture, Educational Technology

1. INTRODUCTION

Student questions offer a direct window into how learning unfolds. They reflect what students find confusing, what they believe matters, and what they feel comfortable admitting they do not understand. In large classes, where instructors and Teaching Assistants (TAs) cannot engage deeply with every student in real time, the questions students ask become a practical signal of where support is needed.

A useful lens for thinking about lecture support is the contrast between *fixed-time* instruction and *self-paced* learning time. Traditional lectures largely follow a fixed schedule, whereas students vary widely in how much time and clarification they require to reach the same level of understanding. Prior work argues for shifting educational experiences toward more flexible time infrastructures to support consistent learning outcomes across students [6, 7, 8]. This framing is often discussed in the context of formative and summative assessments, but it also applies naturally to lecture support. When time is fixed and attention span is limited, students may ask fewer questions, ask different types of questions, or avoid asking altogether; when time is flexible and the interaction is private, students may ask clarification and follow-up questions, watch the lecture in multiple sessions, and pursue understanding iteratively.

At the same time, the conditions under which students ask questions are changing quickly. AI-powered education tools are now widely used alongside traditional course channels such as office hours, discussion forums, and live lecture Q&A. These tools can scale access to help, but they also change basic properties of help-seeking: who students believe is answering their question, whether questions are public or private, how easy it is to ask follow-up questions, and whether students can pause and rewind the material while forming their question. As a result, the questions instructors observe may differ not only because students need help with different concepts, but also because the medium may shape what they choose to ask and how they express it.

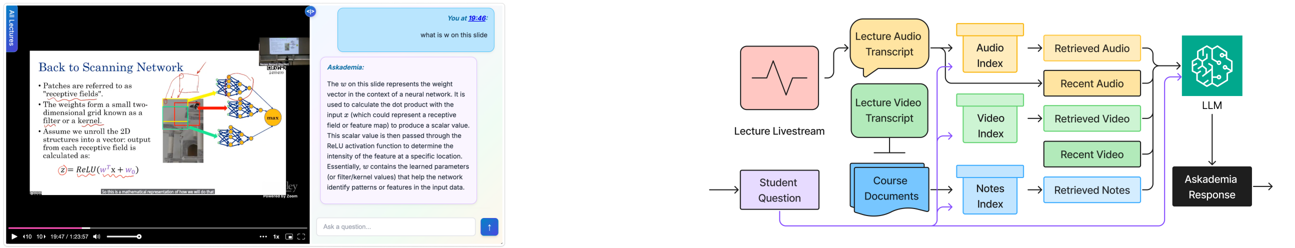
In this paper, we study student information-seeking behavior across two lecture support media with contrasting structures. The first is a synchronous, public live lecture Q&A channel (moderated on Slido[1]), where students post questions that are visible to peers and expect a TA response within the constraints of lecture pacing and a shared queue. The second is Askademia [20], an asynchronous, private AI lecture support tool used while students watch recorded lectures. In this medium, students know an AI system will an-

*Equal contribution.

Meenakshi Mittal, Kaden Tang, Christopher Mach, and Narges Norouzi. The Medium is the Message: How Lecture Q&A Channel Structure Shapes Student Questioning. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyojeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 564–570. International Educational Data Mining Society (2026).

https://doi.org/10.5281/zenodo.21039893

[1] https://www.slido.com/



(a)

(b)

Figure 1: (a) Askademia interface embedded in the lecture experience, enabling grounded question asking at specific lecture moments. (b) Askademia system pipeline for retrieving relevant context and generating grounded responses.

swer them, their questions are private, and they can pause, rewind, and ask multiple follow-up questions.

Askademia [20][2] is a context-aware question-answering tool developed and deployed at UC Berkeley. It processes lecture video and audio to produce searchable, textual representations of spoken utterances and on-screen text, retrieves relevant lecture moments and course notes based on a student question, maintains a sliding window of recent context, and uses a Large Language Model (LLM) to generate a grounded response (Figure 1b). The student-facing interface (Figure 1a) supports questions tied to specific lecture moments and enables rapid back-and-forth as students refine their questions while reviewing material.

We ask essential questions with important implications for both tool design and instructional practice:

RQ1 How do student questions differ between fixed-time, synchronous TA lecture Q&A and self-paced, asynchronous AI lecture support in terms of what students ask and how they express reasoning and confidence?

RQ2 How do question volume, timing, conversation length, and user activity differ across media, and what do these differences suggest about how students use each channel?

To investigate these questions, we introduce a labeling scheme that characterizes questions along three dimensions: (1) intent type (logistical, definition-oriented, analytical, and meta), (2) whether students explicitly or implicitly express their reasoning process, and (3) confidence (neutral, high confidence, and low confidence). We apply this scheme at scale to quantify differences in question content and expression across synchronous and asynchronous media. Finally, we compare question volume, conversation length, and user activity to characterize how question flow differs between live public Q&A and private asynchronous AI support.

2. RELATED WORK

2.1 Student Help-Seeking and Behavior

Student questions are shaped by more than content difficulty: help-seeking is tied to motivation, confidence, and perceived cost (e.g., effort, embarrassment, or fear of asking a "bad" question) [4]. Digital learning tools aim to lower these barriers and increase engagement by changing how and when students can ask for help [2, 3]. As LLM-based tools become more common for student learning support, recent studies emphasize that tool choice and context affect how learners seek help, including whether students externalize uncertainty, ask incremental clarifications, or adopt different strategies [12, 27, 5]. Our work compares how question characteristics shift between a synchronous public medium and an asynchronous private AI medium.

[2] https://www.askademia.org/

2.2 Lecture-Time and Course Q&A Channels

Large courses face a bottleneck: questions arrive continuously, but instructional bandwidth and time are limited [15, 17]. This motivates the design of course Q&A channels to manage scale, redundancy, and timeliness. Tools such as Slido impose distinctive constraints: questions are posted in a shared, public queue, and participation is shaped by lecture pacing and limited response capacity. Prior work in computing education documents both the promise and the practical friction of these at-scale question channels [15, 17]. Our comparison between synchronous, public lecture Q&A and asynchronous, private AI-enabled Q&A examines how these constraints are reflected in the questions students ask and how they participate.

2.3 Retrieval-Augmented and Multimodal AI Course Assistants

Recent educational assistants use LLMs to support student questions in course environments [28, 23, 9]. These systems are often deployed as virtual assistants to reduce logistical overhead and scale students' access to course support [15, 17, 11, 14]. They have become more contextualized and course-aligned using Retrieval-Augmented Generation (RAG) [13, 18, 1]. In parallel, lecture and video tooling have explored how to represent what happened in a lecture through transcripts and visual signals [16]. More recent multimodal methods improve robustness by aligning noisy Automated Speech Recognition (ASR) with visual context [24] and by incorporating video frames into scientific transcription [26]. Broader progress in multimodal and visual question answering further motivates treating lecture content as a multimodal substrate that can be queried [25, 10].

2.4 Analyzing Student Questions

Prior work uses qualitative coding of question intent types and discourse features, paired with reliability checks, as well as automated NLP approaches for categorization at scale. With the growth of LLMs, prompting-based labeling has become a practical way to approximate consistent categorization over large corpora of student discourse [27, 5]. Studies of AI course support also highlight risks around consistency, misalignment with instructional norms, and equity when scaling support [19, 9]. Our paper introduces a compact labeling scheme and applies automated, human-validated labeling to compare patterns across synchronous, public Q&A and asynchronous, private AI Q&A. Rather than evaluating answer quality or system components, our work asks how information-seeking behavior differs when students interact with an AI system privately and asynchronously, compared with a synchronous, public channel.

3. METHODS

3.1 Lecture Support Media

We study student information-seeking behavior in two large upper-division Computer Science (CS) and Data Science (DS) courses at the University of California, Berkeley, with a combined enrollment of around 1,500 students. In both courses, students could ask questions via synchronous and asynchronous channels. In the synchronous medium, students posed questions electronically to TAs during the lecture via Slido. Students contributed to a shared channel of questions, where they could see contributions from other students and TAs. Each lecture lasted approximately 80 minutes, and students could attend live in-person or via Zoom. In the asynchronous medium, students directly asked Askademia while watching recorded lectures on the Askademia platform. Each student contributed to an individual channel of questions, could pause, rewind, and speed up lectures, and could not see other students' conversations.

Per IRB approval for this study,[3] All data included in our analyses were collected and used in accordance with the approved protocol. This protocol covered both sources of student questions considered in our analyses, namely synchronous questions submitted through Slido during lecture and asynchronous questions submitted through Askademia. Our analysis includes **992** synchronous questions (607 CS, 385 DS) asked via Slido across 56 lectures and **4,511** (2,721 CS, 1,790 DS) asynchronous questions from Askademia users who consented to the use of their anonymized questions for research.

3.2 Observable Characteristics

3.2.1 Labeling Scheme

Two annotators conducted exploratory analysis on a random sample of 100 questions and developed a three-dimensional labeling scheme: **Intent Type**, **Reasoning Process**, and **Confidence**. We treat this as a compact, study-specific framework rather than a standardized educational taxonomy. We selected these dimensions because they capture both what students ask and how they ask it, which aligns with our goal of comparing question behavior across synchronous and asynchronous media. The labeling criteria are described below:

[3]This study was approved by the UC Berkeley Institutional Review Board under Protocol ID 2025-09-18906.

Table 1: Representative Examples of Question Labels

Label	Example
	Intent Type
Logistical	"Which textbook are we using?"
Definition	"What is a partial derivative"
Analytical	"Why square the residuals instead of taking an absolute value?"
Meta	"draw me a picture"
	Reasoning Process
Implicit	"Does the denominator of softmax function include the numerator?"
Explicit	"Why does high bias matter? can't we just add in a factor to account for it and try and min the variance?"
	Confidence
Low	"Im confuse, if i have y = mx + b, and i have y(x) = 0, then won't it be 0 = mx +b and thus one solution?"
Neutral	"How is clustering different from classification?"
High	"Bias is the expected difference between predicted and actual though right"

- **Intent Type:** The type of content. Logistical pertains to course operations and policies; Definition pertains to identifying course concepts; Analytical pertains to explaining course concepts in depth; and Meta refers to all questions unrelated to the course. Questions could be labeled as both Analytical and Definition.
- **Reasoning Process:** Whether a student expresses their reasoning process implicitly or explicitly. Explicit questions reveal a student's train of thought, while Implicit questions do not.
- **Confidence:** How confidently a student asks a question. Low-Confidence indicates confusion or skepticism; Neutral questions are asked without uncertainty or over-sureness; High-Confidence questions are typically requests for confirmation and validation.

Two annotators labeled a random sample of 30 questions from both modalities. Initial inter-rater reliability was 60% for intent type, 93% for reasoning process, and 83% for confidence. For each question where there was disagreement, annotators discussed their decisions until an agreement was reached. Afterward, one of the two annotators labeled a random sample of 100 questions.

3.2.2 Automated Labeling

We developed an automated labeling system by prompt engineering GPT-4o [21] for classification. The LLM was provided with a labeling scheme and our set of 30 questions labeled by both annotators. We instructed the LLM to provide an explanation for each generated label, and used hyperparameter values of 0.1 for temperature and top-p, which control the creativity of model responses. When evaluated on our test set of 100 labeled questions, our classifier achieved 81%, 94%, and 83% agreement with human labels for intent type, reasoning process, and confidence, respectively. We used this LLM classifier to label all student questions.

3.2.3 Question Length and Sentiment

We measured length and sentiment as additional characteristics of student questions. To determine sentiment, we used pysentimiento [22], a transformer-based library for sentiment analysis. Each question was classified as Positive, Negative, or Neutral.

3.3 Question Volume and Frequency

We recorded the number of questions per lecture in each medium, the conversation lengths, and the total number of questions asked by each unique user. In the synchronous medium, each question opens a thread where students can post follow-up questions and receive additional TA responses; conversation history is measured as the number of student-TA turns within that thread. In the asynchronous medium, information seeking is continuous in a chat-like interface; we treat a question as a follow-up if it is asked within one minute of a previous question. We selected this one-minute window as a simple heuristic to capture short bursts of back-and-forth clarification in a chat setting. To reduce sensitivity to this choice, we also report a stricter alternative in which follow-up questions are linked only when they share the same lecture timestamp, indicating that a student paused the lecture video to engage further with a specific moment in the recording. We further analyze timestamp-anchored interaction patterns in the asynchronous medium, including clusters of questions asked at the same lecture timestamp and questions asked after rewinding the lecture.

4. RESULTS

4.1 RQ1: Question Content and Expression

We study how question content and expression differ across media by analyzing five characteristics: intent type, reasoning process, confidence, sentiment, and question length.

Table 2: Proportion of each intent type for questions asked in synchronous and asynchronous media, with p-values from a two-proportions z-test. Results are grouped by course.

Intent	Course	Synch	Asynch	Diff	p
Logistical	CS	0.102	0.033	0.069	< 0.001
	DS	0.242	0.064	0.178	< 0.001
Definition	CS	0.245	0.402	-0.157	< 0.001
	DS	0.255	0.368	-0.113	< 0.001
Analytical	CS	0.616	0.513	0.103	< 0.001
	DS	0.475	0.423	0.052	0.062
Meta	CS	0.048	0.072	-0.024	0.029
	DS	0.047	0.153	-0.106	< 0.001

Intent type. Table 2 shows the proportion of each intent type for both media. Questions labeled as both Analytical and Definition contributed to proportion estimates for both categories. For both courses, the proportions of questions that were Logistical and Analytical were higher in the synchronous medium than in the asynchronous one. In comparison, for both courses, the proportions of student questions that were Definition and Meta were significantly lower in the synchronous medium than the asynchronous one.

Reasoning process. Across media, student reasoning patterns were similar. For the CS course, 10.87% of questions expressed explicit reasoning in the synchronous medium compared to 11.44% in the asynchronous one. For the DS course, 8.31% of questions expressed explicit reasoning in the synchronous medium compared to 9.05% in the asynchronous one. These differences were not statistically significant.

Confidence. For the CS course, the proportion of high-confidence questions was lower in the synchronous medium (0.04) than in the asynchronous (0.09), with $p < 0.001$. Similarly, for the DS course, the proportion of high-confidence questions in the synchronous medium (0.021) was roughly half of that in the asynchronous (0.041), although this difference was not statistically significant. For low- and medium-confidence (neutral) questions, differences between media were marginal.

Sentiment. A higher proportion of questions in CS course expressed negative sentiment in the synchronous medium (0.11) compared to the asynchronous (0.075), with $p = 0.004$. However, the proportion of negative questions in the DS course was similar for both media, ranging from 0.119 in the synchronous medium to 0.111 in the asynchronous one. In only the CS course, a lower proportion of questions were neutral in the synchronous medium (0.885) compared to the asynchronous one (0.916) with $p = 0.014$. For positive questions, differences between media were not statistically significant in both courses.

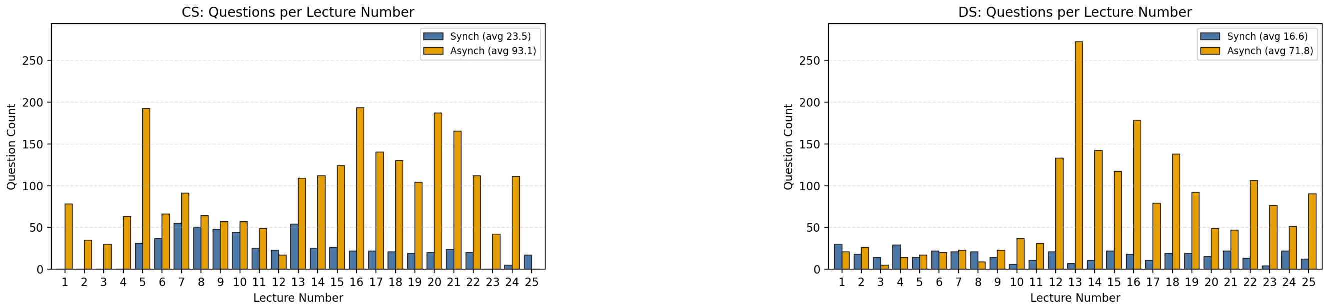
Question length. We report results from a two-sample t-test. For the CS course, the mean question length was slightly greater in the synchronous medium (15.53 words) than in the asynchronous one (14.91 words). For the DS course, mean question length was significantly greater in the synchronous medium (16.15 words) than the asynchronous one (12.86 words), with $p < 0.001$.

4.2 RQ2: Question Volume and Frequency

To study how interaction modality shapes question flow, we compare synchronous in-lecture questions (Synch TA) with asynchronous questions (Asynch Askademia) along three timing/frequency dimensions: per-lecture volume, conversation structure, and per-user participation concentration.

Per-lecture volume. As shown in the Figure 2, Asynch Askademia consistently receives substantially more questions per lecture than Synch TA within each course. In the CS course, the mean questions per lecture was 93.12 (Asynch) vs. 23.52 (Synch), with median questions per lecture 91 vs. 22. In the DS course, mean questions per lecture were 71.84 (Asynch) vs. 16.64 (Synch), with medians 49 vs. 18.

Figure 2 also shows a noticeable mid-semester surge in asynchronous question volume, aligning with the midterm period when students revisit past material. In contrast, Synch TA volume shows a flatter pattern over the same span, with no comparable spike. This divergence suggests that stu-



(a) CS course

(b) DS course

Figure 2: Distribution of questions per lecture, Synch TA vs. Asynch Askademia, shown separately for the CS and DS courses.

dents used the asynchronous medium for extended review and studying, while the live Slido channel may be used more for immediate, in-the-moment clarifications.

Conversation structure. Asynchronous questions are submitted with an associated lecture timestamp, allowing us to characterize how students concentrate questions around particular moments. We define a "paused" conversation as a set of two or more questions, with the same lecture timestamp, from a single student. Across **4,511** asynchronous questions, **2,300** are part of a "paused" conversation (51.0% of all Asynch questions). This behavior has no direct analog in the live TA setting, but highlights a distinct form of demand: students often want to pause at a specific point in the lecture and ask several tightly coupled questions to reach understanding, reinforcing the value of an always-available support channel anchored to fixed points in time.

We also observe **230** "rewind" questions, in which a student asks a question at an earlier lecture timestamp after having asked a question later in the lecture. This further supports the interpretation that asynchronous help-seeking is embedded in non-linear lecture consumption: students pause and move back and forth through the recording, revisiting earlier moments to resolve confusion when it arises.

Figure 3 indicates longer conversations in the asynchronous medium. Mean conversation length is 1.93 (Asynch) vs. 1.06 (Synch). This remains true under a stricter follow-up definition that only links questions with the same lecture timestamp (i.e., "paused" conversations), rather than linking questions within 60s; the asynchronous mean conversation length is still higher (1.54 vs. 1.06), with $p < 0.001$. Asynchronous conversations more often extend beyond single-question interactions, whereas synchronous interactions are predominantly one-shot. This pattern is consistent with students treating the asynchronous tool as a conversational agent: they can ask an initial question, read the response, and then refine or extend their inquiry without worrying about taking up public airtime or interrupting the lecture flow.

Participation concentration. Figure 4 indicates substantially more active users in Asynch Askademia than in the Synch medium. In the Synch medium, only 31 users asked

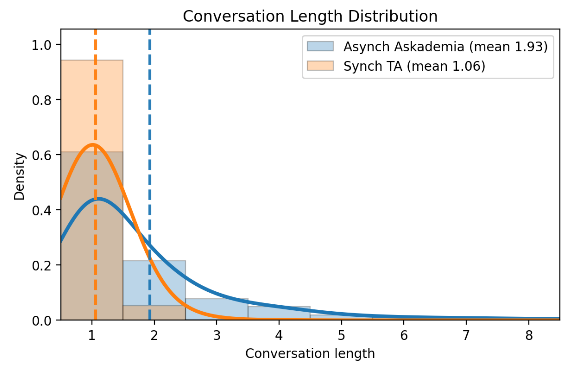


Figure 3: Conversation-length distribution (1 = no follow-up, 2 = one follow-up, etc.)

at least 5 questions throughout the whole semester, whereas in Asynch, 121 users asked at least 5 questions—almost **4**× as many relatively active users. The percentile markers in the figure show the same pattern of heavier engagement in Asynch: Asynch has p90 = 48.0 and p99 = 221.5, compared with Synch, which has p90 = 5.2 and p99 = 27.9. The Synch distribution is visibly concentrated in the lowest bins, suggesting that most students either ask a few questions or rely on the shared channel rather than repeatedly posting. In contrast, Asynch exhibits a much longer right tail, consistent with many more students returning to the tool across the term and using it as a sustained support resource while learning at their own pace. This broader set of active users suggests that a private, self-paced medium may extend meaningful information seeking beyond a small set of highly vocal students who participate in public live Q&A.

5. DISCUSSION

Labeling as a lens on medium differences. Our framework reveals systematic shifts in the types of uncertainty students externalize. The synchronous TA channel contained a larger share of logistical and analytical questions. In contrast, the asynchronous AI channel contained more definition-oriented and meta questions, along with a higher share of high-confidence, confirmation-style questions. One possible interpretation is that students sometimes used the asynchronous tool to validate their understanding of lecture

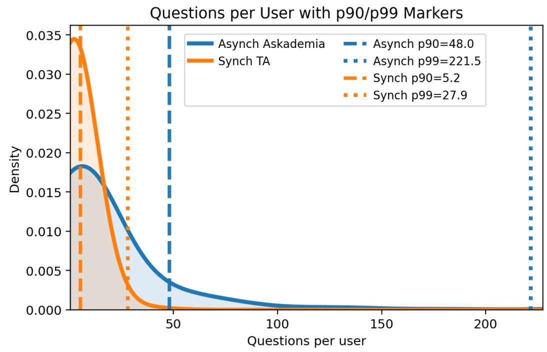


Figure 4: Distribution of the number of questions per user.

concepts rather than only to seek first-pass explanations. Reasoning explicitness remained low in both settings, suggesting that the two media differed more in what students asked than in how they articulated their reasoning. These label distributions suggest that the medium may shape which types of questions are easier to ask, socially appropriate, and likely to receive a useful response.

Participation and timing. Beyond content, the two media differed in scale and participation. The asynchronous medium received substantially more questions per lecture, showed stronger temporal variation, and was used repeatedly by a broader set of students. In contrast, the synchronous TA channel was more concentrated among a smaller group of frequent question askers. While we cannot attribute these differences to a single cause, they are consistent with medium affordances that may reduce barriers to asking, alongside other possible influences such as self-selection into support channels, differences in student expectations, and variation in when students seek help. Though we did not directly test this, shorter asynchronous questions may also align with students using the tool for brief, iterative clarification rather than composing longer, standalone questions. Similarly, some asynchronous questions may have been asked during review rather than first exposure, whereas synchronous questions were asked while students encountered the material live. If so, some observations may reflect not only channel structure but also differences in when students sought help.

Fixed-time instruction, variable-time learning. "Paused" conversation patterns, "rewind" questions, and a higher rate of multi-turn interaction in the asynchronous setting highlight a mode of help-seeking that is difficult to realize in live lecture Q&A. Students appear to pause at specific moments, ask clusters of related questions, and iteratively refine understanding before proceeding. They also sometimes return to earlier lecture moments after moving later in the recording. Unlike the synchronous setting, where questions must be posed under real-time lecture constraints, the asynchronous medium decouples help-seeking from the live flow of instruction. Students can spend variable time at moments of confusion while the instructional timeline remains fixed in the background. This supports a pause-and-repair workflow without the pacing costs of interrupting a live lecture, which may contribute to both longer conversation structure and higher overall demand.

These results motivate treating Q&A channels as part of the instructional environment rather than interchangeable interfaces. Overall, the medium structure was associated with differences in what students asked, who participated, and whether help-seeking took the form of a one-shot request or a more iterative process. Future work should test the generalizability of these patterns and connect medium-specific traces, including confirmation-heavy and "paused" usage, to downstream measures of engagement and learning.

Limitations. Our analysis is limited by the covariates available in the dataset. We did not have access to student-level background measures such as prior GPA, demographics, attendance mode, or prior usage history, nor to course-level covariates such as lecture difficulty or assessment timing. As a result, we could not adjust for potentially important differences in which students used each support channel, and unmodeled confounders, including self-selection and differences in when students chose to seek help, may also contribute to the patterns we observe.

More fundamentally, our comparison does not isolate the effect of any single factor. The two channels differ simultaneously along structural dimensions such as synchronicity, privacy, responder identity, and lecture-viewing context, as well as harder-to-measure sociotechnical ones such as students' familiarity with and acceptance of AI tools, their expectations about response quality and authority, their perceptions of the AI's institutional role, and the social costs of asking. The observed differences may therefore reflect a bundle of channel affordances, interaction norms, and user expectations rather than the independent effect of the interaction medium alone. Our findings are also drawn from two computing courses in one institutional context, which may limit generalizability. Future work with richer student- and course-level data, and designs that vary these features more independently, could better disentangle these explanations.

6. CONCLUSION

Our findings suggest that the transition from synchronous, human-led Q&A to asynchronous, AI-driven support is associated with a shift from fixed-time/variable-learning to flexible-time/fixed-learning. In traditional synchronous lectures, the rigid temporal structure may function as a bottleneck, placing students in a competitive, public environment where learning remains "variable" for those who cannot keep pace with the live stream. In contrast, the asynchronous medium supports "flexible time" when inquiry is decoupled from the lecture's live delivery, allowing students to treat conceptual mastery as a "fixed" goal. The asynchronous medium is associated with students engaging in denser, more iterative, and definition-oriented dialogues, which may reflect lower social costs and fewer time constraints. These patterns suggest that medium structure may support a more equitable instructional model, in which time spent is the variable and depth of understanding is the constant.

7. REFERENCES

[1] H. A. Alawwad, A. Alhothali, U. Naseem, A. Alkhathlan, and A. Jamal. Enhancing textbook question answering task with large language models and retrieval augmented generation, 2024.

[2] A. Ali. Using padlet as a pedagogical tool. *Journal of Learning Development in Higher Education*, 2021.

[3] J.-M. Campillo-Ferrer, P. Miralles-Martínez, and R. Sánchez-Ibáñez. Gamification in higher education: Impact on student motivation and the acquisition of social and civic key competencies. *Sustainability*, 2020.

[4] Y. F. Cheong, F. Pajares, and P. S. Oberman. Motivation and academic help-seeking in high school computer science. *Computer Science Education*, 2004.

[5] W. Gan, Z. Qi, J. Wu, and J. C.-W. Lin. Large language models in education: Vision and opportunities. In *2023 IEEE International Conference on Big Data (BigData)*, 2023.

[6] D. Garcia, A. Fox, S. Russell, E. Ambrosio, N. Terrell, M. Silva, M. West, C. Zilles, and F. Shakir. A's for all (as time and interest allow). In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, volume 1, 2023.

[7] D. Garcia, C. McMahon, Y. Garcia, M. West, and C. B. Zilles. Achieving "a's for all (as time and interest allow)". In *Proceedings of the Ninth ACM Conference on Learning @ Scale*, 2022.

[8] D. Garcia, C. McMahon, Y. Garcia, M. West, and C. B. Zilles. Software support for "a's for all". In *Proceedings of the Ninth ACM Conference on Learning @ Scale*, 2022.

[9] Y. Hicke, A. Agarwal, Q. Ma, and P. Denny. Ai-ta: Towards an intelligent question-answer teaching assistant using open-source llms, 2023.

[10] M. F. Ishmam, M. S. H. Shovon, M. F. Mridha, and N. Dey. From image to language: A critical analysis of visual question answering (vqa) approaches, challenges, and opportunities. *Information Fusion*, 106, 2024.

[11] M. Kazemitabaar, R. Ye, X. Wang, A. Z. Henley, P. Denny, M. Craig, and T. Grossman. Codeaid: Evaluating a classroom deployment of an llm-based programming assistant that balances student and educator needs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024.

[12] H. Kumar, M. Reza, J. Mitchell, I. Musabirov, L. Zhang, and M. Liut. Understanding help-seeking behavior of students using llms vs. web search for writing sql queries, 2024.

[13] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. Retrieval-augmented generation for knowledge intensive nlp tasks. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[14] R. Liu, C. Zenke, C. Liu, A. Holmes, P. Thornton, and D. J. Malan. Teaching cs50 with ai: Leveraging generative artificial intelligence in computer science education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, volume 1, 2024.

[15] T. M. MacWilliam and D. J. Malan. Scaling office hours: managing live q&a in large courses. *Journal of Computing Sciences in Colleges*, 28, 2013.

[16] C. Mahipal, L. Angrave, Y. Xie, B. Chatterjee, H. Wang, and Z. Qian. "what did i just miss?!" presenting classtranscribe, an automated live-captioning and text-searchable lecture video system, and related pedagogical best practices. In *Proceedings of the 2019 ASEE Annual Conference & Exposition*, 2019.

[17] S. Mirhosseini, A. Z. Henley, and C. Parnin. What is your biggest pain point?: An investigation of cs instructor obstacles, workarounds, and desires. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, volume 1, 2023.

[18] C. Mitra, M. Miroyan, R. Jain, V. Kumud, G. Ranade, and N. Norouzi. Retllm-e: Retrieval-prompt strategy for question-answering on student discussion forums. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2024.

[19] M. Mittal, A. Bailey, V. Phelps, M. Miroyan, C. Mitra, R. Jain, R. Niousha, G. Ranade, and N. Norouzi. Raising the bar: Automating consistent and equitable student support with llms. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2*, volume 2, 2025.

[20] M. Mittal, G. Tyagi, A. Bailey, G. Ranade, and N. Norouzi. Askademia: A real-time ai system for automatic responses to student questions. In *Artificial Intelligence in Education*, 2025.

[21] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, et al. Gpt-4 technical report, 2023.

[22] J. M. Pérez, M. Rajngewerc, J. C. Giudici, D. A. Furman, F. Luque, L. Alonso Alemany, and M. V. Martínez. pysentimiento: A python toolkit for opinion mining and social nlp tasks, 2021.

[23] L. Qi, J. D. Zamfirescu-Pereira, T. Kim, B. Hartmann, J. DeNero, and N. Norouzi. A knowledge-component-based methodology for evaluating ai assistants, 2024.

[24] N. Shvetsova, A. Kukleva, X. Hong, C. Rupprecht, B. Schiele, and H. Kuehne. Howtocaption: Prompting llms to transform video annotations at scale. In *Computer Vision – ECCV 2024*, 2024.

[25] S. Uppal, S. Bhagat, D. Hazarika, N. Majumder, S. Poria, R. Zimmermann, and A. Zadeh. Multimodal research in vision and language: A review of current and emerging trends. *Information Fusion*, 77, 2022.

[26] M. Wang, Y. Wang, T.-T. Vu, E. Shareghi, and G. Haffari. Exploring the potential of multimodal llm with knowledge-intensive multimodal asr. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024.

[27] S. Wang, T. Xu, H. Li, C. Zhang, J. Liang, J. Tang, P. S. Yu, and Q. Wen. Large language models for education: A survey and outlook, 2024.

[28] J. D. Zamfirescu-Pereira, L. Qi, B. Hartmann, J. DeNero, and N. Norouzi. 61a bot report: Ai assistants in cs1 save students homework time and reduce demands on staff. (now what?). In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, volume 1, 2025.