# ClassBank: A Comprehensive Resource for Classroom Discourse Analysis in Education

**Shireen Al-Adeimi**[a] **and Catherine Snow**[b], [a] Michigan State University, East Lansing, MI, United States; and [b] Harvard Graduate School of Education, Cambridge, MA, United States





## Key Points

- Emerging from the Child Language Data Exchange System (CHILDES), TalkBank is a comprehensive digital repository designed for sharing and analyzing spoken language across contexts.
- ClassBank is a specialized sub-database within TalkBank focusing on classroom discourse, and offers researchers tools to transcribe, annotate, and analyze interactions in educational settings using multimodal data.
- Analytic tools for discourse analysis include Collaborative Commentary, which enables group-based data coding and analysis.
- A large, multi-institution collaboration focused on Academically Productive Talk (APT) utilized ClassBank to code K-12 classroom discussions using three instruments—the Low-Inference Discourse Observation tool, Quality Talk, and the Science Discourse Instrument—demonstrating ClassBank's capacity to support diverse analytical approaches.
- Integration of automatic speech recognition and involving educators, students, and community members through citizen science projects will further expand ClassBank's database and promote inclusive research practices.

## Abstract

ClassBank is a sub-database within the TalkBank infrastructure and serves as a key resource for researchers and educators studying classroom discourse. This article provides an overview of ClassBank's tools for capturing, transcribing, and analyzing classroom interactions, highlighting its strengths in multimodal analysis and collaborative coding. It also discusses the use of ClassBank in the Academically Productive Talk project to demonstrate its flexibility in supporting various analytic approaches. The article concludes by considering additional uses of ClassBank and its future applications in research and practice.

## Introduction

Classroom discourse plays a crucial role in shaping how students construct knowledge, language, and critical thinking skills (Alexander, 2008; Murphy et al., 2009; Snow et al., 2009). By participating in dialogic, academically productive discussions with peers and teachers, students learn to articulate their ideas, engage in argumentation, and collaborate in the meaning-making process. For educational researchers, analyzing these interactions offers valuable insights into teaching practices that best support students' learning, as well as the range and quality of student engagement, and how they relate to learning outcomes across diverse contexts. However, studying classroom discourse presents unique challenges due to its complexity, as it requires versatile tools that can capture both the verbal and non-verbal elements of communication. Traditional methods of discourse analysis are often labor-intensive, highlighting the need for an efficient platform that supports data collection, transcription, coding, and annotation. ClassBank, a specialized sub-database within the broader TalkBank infrastructure, addresses this need by offering a platform for the study of classroom talk. It allows researchers to explore a wide range of research questions related to language development, pedagogy, and educational equity. By integrating advanced analytic tools, ClassBank facilitates the close examination of classroom

interactions, thereby supporting large-scale research efforts and cross-study comparisons that contribute to deeper understanding of educational practices.

## Historical Background

In 1984 Brian MacWhinney and Catherine Snow, with funding from the MacArthur Foundation, launched the Child Language Data Exchange System (CHILDES), a digital archive of transcripts of parent-child interactions, in an effort to support research on child language acquisition by expanding the value of laboriously produced transcripts through data-sharing. Promoting the desired collaborative data-sharing required establishing a standardized set of transcription conventions (Codes for the Human Analysis of Transcripts, CHAT), and was incentivized by the development of digital analysis tools (Computerized Language ANalysis, CLAN). CHILDES was the source of data used in multiple research publications even during its first few years of existence (MacWhinney & Snow, 1990). The generosity of Roger Brown in distributing mimeographed copies of the transcripts of his participants Adam, Eve, and Sarah to students and colleagues during the 1960s was the model for CHILDES.

Although the CHILDES database was for the first several years of its existence primarily populated by transcripts of English-speaking parents with their preschool-aged children, it quickly expanded to include child language and parent-child interaction data from children learning languages other than English, from bilingual children, and from children with various developmental disabilities, as well as to include data from child narrations or narrative retellings, data not collected in naturalistic interactions. By 2000 the growing interest in classroom discourse, ranging from early childhood settings such as those studied by Grøver (e.g., Aukrust, 2007; Grøver et al., 2022) to fourth and eighth grade classroom videos from the TIMMS cross-national study (Stigler, Gallimore, & Hiebert, 2000) to discussion-based classrooms at university (e.g., Koschmann & LeBaron, 2002) led to the establishment of ClassBank in the now renamed database structure TalkBank.

## Uses and Features of ClassBank

ClassBank, a specialized component of the broader, open-source repository TalkBank, serves as an important resource for examining classroom discourse within educational settings. Its tools enable researchers to capture, transcribe, and annotate a wide range of classroom interactions in K-12 and higher education contexts, making it particularly valuable for projects investigating how talk supports learning and other outcomes. TalkBank has contributed to thousands of research publications and has practical applications in fields such as language therapy, clinical diagnosis, and second language education (MacWhinney, 2019).

ClassBank's strengths lie in its ability to integrate multimodal data—audio, video, and transcripts—and its flexibility in accommodating various analytic approaches (MacWhinney, 2019). This multimodal data integration allows video recordings to be linked directly to transcripts, which gives researchers the ability to analyze both verbal and nonverbal elements of classroom interactions. Researchers can observe non-textual aspects of interactions, such as the physical positioning of interlocutors, gestures, and eye contact, alongside transcriptions of classroom talk. Such a multimodal view of communication enables deeper insights into how teachers and students co-construct meaning during discussion, thereby contributing to a more holistic analysis of classroom dynamics.

Key principles of TalkBank are "maximally open data sharing" and "interoperability" (MacWhinney, 2019, p. 1921). These principles ensure that data collected and analyzed through ClassBank are easily accessible and reusable by researchers across institutions, which fosters large-scale collaborations across institutions and cross-study comparisons. Moreover, the platform's flexibility supports educational research in diverse linguistic, cultural, and pedagogical contexts.

One of ClassBank's most notable features is its Collaborative Commentary (CC) system, which enables researchers to annotate transcripts collaboratively and comment on specific moments in the transcript (MacWhinney & Fromm, 2023). Through this feature, researchers can comment on transcripts either individually or in groups, and select from predefined codes created by the group, which enables consistent coding and ease of collaboration across projects (MacWhinney & Fromm, 2023). This feature is especially useful in educational research where interrater reliability is essential for coding systems, as researchers seek to develop a shared understanding of coding conventions and interpretations. Researchers using ClassBank can apply various coding schemes to the data, using systems like CHAT or creating their own custom codes (MacWhinney & Wagner, 2010). This flexibility allows for diverse methodologies to be applied to the same data, making ClassBank a versatile and multifaceted tool in the field of discourse analysis.

## ClassBank's Role in Research and Practice

The versatility of ClassBank extends beyond its utility for academic research; it is also a valuable resource for teacher education and professional development. Teachers and pre-service educators can review recorded classroom interactions to reflect on their discourse practices, critically examining how they ask questions, provide feedback, and scaffold student contributions. By engaging with multimodal data, educators can closely examine how their verbal and non-verbal communication supports—or hinders—student engagement and learning. This reflective practice is central to the development of effective teaching practices, particularly those that foster dialogic learning, where students are encouraged to build on each other's ideas and engage in meaningful, academically productive discussions.

Additionally, ClassBank's potential for examining and improving educational equity is significant. Researchers can use the platform to study how discourse practices in classrooms vary across different demographic groups, examining patterns of participation, engagement, and access to learning opportunities. For example, ClassBank can be used to analyze whether certain groups of students—such as bi/multilingual students—participate as fully in classroom discussions as their peers. This type of analysis can inform interventions aimed at promoting more inclusive discourse practices, to ensure that all students have the opportunity to contribute meaningfully to classroom conversations.

## The Academically Productive Talk (APT) Project

ClassBank's wide-ranging applicability makes it a valuable tool for both researchers and practitioners. In this section, we examine a project utilizing ClassBank's platform and analytic tools to analyze classroom discussions across multiple subject areas in K-12 contexts, to shed light on how teachers and students engage in productive, dialogic talk that fosters linguistic and academic skills.

The Academically Productive Talk (APT) project is a multi-institutional collaboration funded by the Spencer Foundation, with aims to guide researchers and practitioners in understanding the nature and the role of classroom discussions across different grades and subject areas. The project's website (linked below) highlights curated and annotated video segments of classroom discussions, accompanied by commentary from the research team noting the academically productive features of the talk observed in each video.

To enable researchers to examine various aspects of the talk itself, the APT team utilized the ClassBank platform to transcribe and code all segmented videos using three distinct instruments: the Low-Inference Discourse Observation (LIDO) tool (Al-Adeimi & O'Connor, 2017; LaRusso et al., 2023), Quality Talk (QT; Murphy & Firetto, 2017), and the Science Discourse Instrument (SDI; Fishman et al., 2017). Each of these tools offers a unique approach to analyzing classroom discourse, providing complementary insights into different facets of classroom talk.

The LIDO tool captures the presence and form of teacher and student talk, each of which is categorized along a spectrum of dialogicity. Using CHAT conventions, classroom discussions analyzed with the LIDO are coded at the speaker turn level. This detailed coding enables researchers to observe teacher question types (open, semi-open, or closed), scaffolding moves (e.g., prompting student-to-student engagement and pressing for reasoning) as well as student responses (e.g., student-to-student talk, student questions, or minimal responses). Together, these codes offer insights about the range of talk moves produced by teachers and students, as well as their potential for dialogicity.

On the other hand, the SDI assesses both the presence and quality of talk along three student and three teacher discursive forms, which are provided at the end of each transcript in ClassBank. For example, the tool assigns a value (from 0—4) indicating the frequency and nature of teacher questions, with 1 indicating a closed-ended question and 4 indicating consistent use of open-ended questions. Similarly, the quality of teacher and student discursive forms are evaluated as emerging or proficient, thereby providing a holistic view of the productivity of classroom discussions.

Finally, QT focuses on the types of questions and responses exchanged between students and teachers. The instrument uses CHAT "gem markers" to signify the beginning and end of each codable event within the transcript, with codes like authentic or uptake questions, as well as *exploratory* or *cumulative* responses that indicate the level of student engagement in a given exchange. By concentrating on the dialogic features of questions and responses, QT helps identify how certain types of teacher and student exchange foster deeper understanding of texts and other content.

Each tool's distinctive coding scheme allows for a multifaceted analysis of classroom talk, and offers insights into both the structure and content of productive discussions. By applying these instruments through the ClassBank platform, the team studying academically productive talk was able to capture a comprehensive understanding of how such talk manifests across different classrooms and subject areas. Furthermore, utilizing ClassBank allows transcripts to be linked to video segments and enables the close and efficient examination of coded transcripts with analytic commands such as frequency (freq), key word and line (kwal), mean length of turn (mlt) (MacWhinney & Wagner, 2010). As such, ClassBank played a crucial role in this research, providing tools to capture, transcribe, and analyze classroom discourse in a way that was both flexible and rigorous.

## Future Directions and Extension of ClassBank

A promising future direction for ClassBank is the integration of automatic speech recognition (ASR) technology, which would allow for automatic generation of transcripts from classroom audio and video recordings (MacWhinney & Fromm, 2023). This could significantly streamline the data analysis process, making it easier for researchers to analyze classroom talk at scale. The integration of ASR, along with ClassBank's existing multimodal data capabilities, broadens its cross-disciplinary potential. Researchers in the fields of psychology, computational linguistics, and cognitive science could use ClassBank data to examine intersections between language processing, social interaction, and cognitive development within classroom contexts.

Additionally, the platform's adaptability opens up possibilities for "citizen science" projects, where teachers, students, and community members can participate in the transcription and coding of classroom interactions (MacWhinney & Fromm, 2023). By engaging a broader community in research, ClassBank can expand its database while fostering deeper and more inclusive pedagogical approaches. Integrating ClassBank into teacher education and professional development programs could also enable educators to engage with research findings directly, which could potentially inform the use of evidence-based practices in classroom settings.

## Conclusion

ClassBank is an invaluable resource for educational researchers and practitioners, and offeres a flexible, multimodal platform for analyzing classroom discourse. Emerging from the TalkBank infrastructure, it has made significant contributions to understanding classroom interactions, discourse practices, and educational outcomes. The use of ClassBank in the APT project demonstrates its capacity to support diverse analytical frameworks, thereby enhancing our understanding of how dialog fosters student learning. With future developments like automatic speech recognition and broader community involvement, ClassBank is poised to further advance research and practice in education.

## References

Al-Adeimi, S., & O'Connor, C. (2017). *Low-inference discourse observation (LIDO) coding manual*. Boston, MA: Boston University.

Alexander, R. J (2008). *Towards Dialogic Teaching: Rethinking classroom talk* (4th ed.). Dialogos.

Aukrust, V. G. (2007). Young children acquiring 2nd language vocabulary in preschool group-time: Does amount, diversity, and discourse complexity of teacher talk matter? *Journal of Research in Childhood Education, 22*(1), 17—37. https://doi.org/10.1080/02568540709594610

Fishman, E. J., Borko, H., Osborne, J., Gomez, F., Rafanelli, S., Reigh, E., & Berson, E. (2017). A practice-based professional development program to support scientific argumentation from evidence in the elementary classroom. *Journal of Science Teacher Education, 28*(3), 222—249. https://doi.org/10.1080/1046560X.2017.1302727

Grøver, V., Rydland, V., Gustafsson, J.-E., & Snow, C. E. (2022). Do teacher talk features mediate the effects of shared reading on preschool children's second-language development? *Early Childhood Research Quarterly, 61*, 118—131. https://doi.org/10.1016/j.ecresq.2022.06.002

Koschmann, T., & LeBaron, C. (2002). Learner articulation as interactional achievement: Studying the conversation of gesture. *Cognition and Instruction, 20*(2), 249—282. http://www.jstor.org/stable/3233875.

LaRusso, M., Al-Adeimi, S., Harbaugh, A., Kim, H. Y., & O'Connor, C. (2023). Capturing classroom talk with the Low inference discourse observation tool: A validation study. *Journal of Early Adolescence, 0*(0), 1—31. https://doi.org/10.1177/02724316231182291

MacWhinney, B. (2019). Understanding spoken language through TalkBank. *Behavior Research Methods, 51*(4), 1919—1927. https://doi.org/10.3758/s13428-018-1174-9

MacWhinney, B., & Fromm, D. (2023). Collaborative commentary for understanding communication disorders. *American Journal of Speech-Language Pathology, 32*, 2580—2588. https://doi.org/10.1044/2023_AJSLP-22-00385

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17*(2), 457—472.

MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung: Online-Zeitschrift zur verbalen Interaktion, 11*, 154—173.

Murphy, P. K., & Firetto, C. M. (2017). Quality talk: A blueprint for productive talk. In P. K. Murphy (Ed.), *Classroom discussions in education* (1st ed., pp. 101—133). Routledge.

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*, 740—764. https://doi.org/10.1037/a0015576

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2*(4), 325—344. https://doi.org/10.1080/19345740903167042

Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS and TIMSS-R video studies. *Educational Psychologist, 35*, 87—100.

## Relevant Websites

APT website: https://www.academically-productive-talk.org/.
ClassBank: https://class.talkbank.org/.
TalkBank: https://talkbank.org/.