Developing an AI-Supported Approach to Identify Instructional Groupings in Early Childhood Education Classrooms

Technical White Paper

December 2025

Prepared by:
Aravind Sundaresan, Leigh Ann DeLyser, Gullnar Syed, Sarah Gerard, and John Niekrasz

SRI

Suggested citation:
Sundaresan, A., DeLyser, L. A., Syed, G., Gerard, S., & Niekrasz, J. (2025). *Developing an AI-Supported Approach to Identify Instructional Groupings in Early Childhood Education Classrooms* [Technical report]. SRI.

Contents

Abstract
A high-quality early childhood classroom is alive with activity. Children may be rotating from whole-group morning meeting to free play centers and then to small-group math instruction, all before lunch. Pre-K teachers looking to make the most of these instructional groupings—and those who support them, such as instructional coaches—can benefit from understanding how classroom time is spent to best focus instructional goals. Many early childhood providers use video to support teacher development, classroom observations, or even opportunities for parents to look in on their children during the school day. Researchers from SRI, supported by the Gates Foundation, explored opportunities to leverage AI and early childhood classroom videos to develop foundational frameworks and AI-supported approaches that promote high quality teaching and learning. After interviews with curriculum providers, early childhood education (ECE) leaders and educators, and funders, the research team selected instructional groupings as a fundamental building block that could be used to support coaching, curriculum implementation at scale, and data for ECE providers for review of implementation. This technical white paper describes the research team's approach to automatically identifying instructional groupings, the data used, and initial performance of tested models on the dataset. We close with a discussion of future implications of this work.

Introduction
The task explored in this white paper is the automatic classification of preschool classroom video segments according to observed instructional groupings such as Whole Group, Small Group, Centers and Free Play, and Independent work (see Table 1 for the full list). The research team considered two approaches to perform this classification.

(1) **Analysis of videos using a suite of pre-trained machine learning (ML) models** for person detection and tracking [1], [2], pose estimation, and activity classification. These methods can characterize a video segment based on the number of people in the video, their body poses, and activity. A rule-based approach could then be used to classify the video segment based on the instructional grouping. Alternatively, a shallow neural network (NN) could be trained to classify the video segment if sufficient training examples are available.

(2) **Analysis of videos using multimodal large language models (MLLMs)** [3] and prompts that describe the instructional grouping rubric. This approach relies on the capabilities of an MLLM to summarize images or videos as text and analyze the text based on user prompts. Contrasting with (1), this approach requires no training data for model supervision or fine-tuning.

The ML suite approach requires integrating various modules to extract features from video and audio (such as person bounding boxes, poses, or activities). This is followed by either training a shallow NN layer to fuse the *features* extracted from video and audio or creating a set of rules to perform classification based on the instructional grouping rubric and the extracted features. The MLLM approach is more general and requires much less development effort as the different classes can be described using text prompts and would be more accessible to the larger early childhood education ecosystem. However, the MLLM approach may not be able to analyze certain aspects in detail, such as the pose of teacher and student. For this work, we determined that the multimodal LLM would be a better fit because it is faster to develop, is more general, and does not require further training and a suitable dataset to train on.

We apply the MLLM approach in this white paper to 13 hours of preschool classroom video data using models such as Google Gemini [5] and the OpenAI GPT [6] family of models. While either model could be used for the video analysis, we used only OpenAI GPT models for evaluation on the dataset described in this white paper. SRI has an Enterprise license for GPT, which ensures that images uploaded into the model will not be used to train models outside of SRI's closed system. Given that images of teachers and students would be included in the classroom images, we selected the highest security model to protect their data privacy.

Instructional Grouping Framework Development

Dataset

Video sequences were collected from four preschool classrooms and totaled 13 hours, with an average of 3 hours and 15 minutes of footage from each classroom. The videos represented multiple instructional moments, such as teacher-led small-group math instruction, as well as general classroom activities such as morning meeting, lunch, and transitions. Videos often contained multiple students moving around the classroom, students receiving one-on-one behavior support, and time spent in centers such as the blocks area, library, or dramatic play kitchen area. In many classrooms, 2–3 educators (often a lead teacher, co-teacher, and/or paraeducator) are leading or monitoring groups of children in concurrent activities.

Team members collected videos using an iPad on a Swivl, a robotic camera tripod mount that rotates 360 degrees to track a person wearing a microphone marker. The lead teacher and assistant teacher both wore lanyard microphone markers clipped to their collars to record audio. The tripod was assembled so that the iPad was at chest height with the teacher and the Swivl rotated to track the movements of the lead teacher around the classroom. This allowed for the field of view to include the lead teacher's interactions with students and the classroom area adjacent to the teacher. Data collectors were present during the recording and did not move the tripod during the recording, unless the device became an obstacle for classroom movements. In these few cases, recordings were stopped, and the camera was repositioned. The research team obtained permission from teachers for the video to be used for this purpose, and parents had the opportunity to "opt out" their child from participation. The dataset will not be released for public use to ensure privacy of participating teachers and children.

A second dataset, containing an additional seven hours of video from 16 classrooms, was added to increase the variety of classroom settings in the existing dataset. These recordings captured a range of classroom activities, with an emphasis on small-group interactions. The research team established a data-sharing agreement with an external researcher to obtain permission to access these videos, which included releases from both teachers and families allowing their use for research purposes [4]. In total, the combined dataset included 20 hours of video from 20 classrooms.

Framework

Two early childhood experts on the team reviewed early childhood classroom observation tools (e.g., COP/TOP [7] [8], CLASS [9], ECERS [10]) to identify approaches to coding instructional groupings. They synthesized and modified these elements and generated an annotation framework to classify video segments for the type of instructional grouping shown in classroom video recordings. Early childhood teachers use a variety of instructional groupings for different purposes. For instance, the teacher may introduce a new concept (data analysis with a basic block graph) during morning meeting by asking children to select their favorite food from three options—pizza, cheeseburger, or mac and cheese—and then graphing responses on the whiteboard. Next, the children move into small groups and center times, in which one teacher works with a small group of children to practice block graphs with new data (e.g., favorite animals and favorite vehicles) for 10 minutes. Concurrently, the remaining children are in centers—some working with Unifix cubes to match cube quantities on a poster, some discussing favorite foods in the dramatic play kitchen, and some perusing books in the library center focused on counting and comparing objects. Many of these activities can look similar to an outside observer. Distinguishing between the groupings can be helpful for an instructional coach or program director looking to see how a classroom spends its time, and if children are spending sufficient time in groupings such as small groups, when high-impact instruction is often best delivered. Identifying the groupings requires considering the teacher activity and the teacher-student interactions and assessing what is happening across multiple student learning groups.

Thus, the annotation framework included definitions of each grouping and precise start and end cues for each grouping. The framework was iteratively modified following practice annotation to refine and tighten definitions as well as to collapse codes that were too similar or could not be accurately distinguished. For instance, pre-K students commonly engage in independent work, either directly with a teacher (one-on-one) or individually completing activities with a teacher monitoring. Originally, independent work was split into three subcodes: one-on-one, independent work at tables, and independent work *not* at tables. After practice coding, we removed the subcode of independent work *not* at tables because this code's definition was too similar to the definition of centers/free play and would be frequently mis-annotated. The original framework also included separate subcodes for meal times (breakfast, lunch, and snack). However, after discussion, the subcodes were determined to be difficult to distinguish and less important than the larger identification that meal time was occurring. The last modification occurred at the suggestion of the development team: Rather than identifying rest time, recess, or no people visible as subcodes to Other, these codes were brought to the level of primary codes. The final framework contains 10 codes, each with a three-letter abbreviation for annotations (Table 1). Three codes are broken down into subcodes. See Appendix A for the complete codebook.

Table 1. Framework and codes

Code	Definition	Abbreviation	Applicable subcodes (abbreviation)
Whole Group	Students standing or seated near teacher in a circle or non-circle	WGR	Circle time (CIR); Carpet time (CAR)
Small Group	Students engaged in the same activity in different areas in the classroom with at least 1 teacher near each group	SGR	-
Centers/Free Play	Students in different areas engaged in different activities, with or without other students	CFP	-
Independent	Teacher working with an individual student or students completing an activity in parallel	IND	One-on-one (UNO), independent work at tables (TAB)
Transitions	Students moving from one instructional grouping to another	TRN	Within classroom (INS), to outside (OUT), to inside (INS)
Meal Time	In class breakfast, lunch, or snack	MEL	-
Rest Time	Students lying down in different areas of the classroom	RST	-
Outside/Recess	Students are outside	REC	-
No People Visible	No students/teachers visible	NOP	-
Other	Anything that does not fit in other groupings	OTH	-

Annotations

Following framework development, the research team created a template spreadsheet for annotating classroom videos (Figure 1). The template allowed annotators to identify the start and end times for episodes of up to three distinct instructional groupings (and their subcodes) occurring concurrently. As an annotator watched a video, they entered the primary grouping (code1) during a time period, and if applicable, any additional groupings co-occurring (code2 and code3). Primary groupings were the grouping that most students (~75%) were engaged with. For example, Figure 1 shows that from 2:00 to 3:30, students were transitioning into the classroom. At 3:31, most students were still transitioning into the classroom, but some students began independent work at tables. Such gradual transitions are typical in early childhood classrooms and can minimize child wait times and maximize learning time. (For example, instead of waiting for all children to enter the classroom and sit down to begin independent work, children begin as soon as they reach their seat, allowing for nearly 2 minutes of learning time that otherwise would have been wait time.) This approach of annotating continuous blocks of classroom video was necessary as instructional groupings and transitions tend to be in larger chunks of time (e.g., 3–20 minutes). Human experts can more efficiently detect transitions in this manner rather than reviewing short snippets of video for labeling. Annotators tested and refined the template with practice videos. The development team used completed spreadsheets containing timestamps and codes for all instructional grouping segments to represent correct “ground-truth” information during system evaluation.

Figure 1. Instructional grouping annotation

A	B	C	D	E	F	G	H	I
start	**end**	**code1**	**sub1**	**code2**	**sub2**	**code3**	**sub3**	**Duration**
0:02:00	0:03:30	trn	int					0:01:30
0:03:31	0:05:20	trn	int	ind	tab			0:01:49
0:05:21	0:05:50	cfp		ind	tab	trn	int	0:00:29
0:05:51	0:09:32	ind	tab	trn	int			0:03:41
0:09:33	0:34:00	cfp		ind	tab			0:24:27
0:34:01	0:39:06	cfp						0:05:05
0:39:07	0:40:41	trn	ins					0:01:34

Annotator training and reliability

Three human annotators with experience in early childhood classrooms attended a 1.5-hr training and practiced coding six 10-minute video segments. The annotator team compared their annotations to those of the lead annotator, and annotators were given feedback to refine their codes to align with the lead annotator. Once the reliability threshold of 80% exact agreement on both content and timestamp on the training videos was achieved, annotators coded the remaining video data across 2 weeks. For ease of annotation, videos assigned to annotators were portions of longer recordings, approximately 20 minutes each. Annotators were instructed to watch the video at 1- or 2-times speed and note the start time, end time, and instructional grouping(s) present. 10 segments were coded by two annotators and compared for reliability. Annotations were first split into common time segments (with the same start time and end time) to allow for meaningful comparison. The agreement score for each common segment was the Intersection-over-Union (IoU) of the set of annotations from each annotator. If the annotators agreed perfectly, the score was 1. If there was no common annotation, the score was 0. The total score was computed by weighting the score of each segment by its duration. Annotators achieved 91.6% in weighted IoU. Exact agreement was also calculated between annotators and weighed by timestamp duration length; 83.5% of annotations were matched exactly, demonstrating high interrater agreement.

MLLM-Based Video Classification and Evaluation

We used MLLMs to classify video segments independently as described in the steps below. In these models, the video segments are obtained by splitting the original video sequences based on the timestamps obtained from human annotations. The system does not segment the video sequence automatically.

(1) The video segment is sampled every 2 seconds to obtain a list of image frames. These frames are resized to a smaller resolution that is sufficient for observing the activities in the video while minimizing the image and token size. It is usually not necessary to sample more frequently to analyze the activity in the video.

(2) These frames are encoded and processed by the MLLM using the Python API for GPT.

The instructions to the MLLM are specified in Figure 2.

(3) The MLLM is instructed to return the results in JSON format by providing a schema. The results include the instructional grouping class, score, and a brief text description of the instruction. The results include a reason for the class chosen and the frames coded to promote transparency for the decision-making of the MLLM in response to a thematic need for trust from participants interviewed by the project.

Figure 2 displays an evaluation results table showing side-by-side comparison of ground truth with MLLM-generated predictions. Columns 2–4 indicate human annotations. The far right columns show the output of two models for comparison. The cell background color indicates if the model classification is correct. If the primary category predicted by the model matches any of the three human annotated codes, the classification is correct (blue). For instance, in Figure 2, gpt-5-mini predicts TRN as the primary category for Segment 000. The ground truth is also TRN, so we label this cell as correct. Segment 001, however, is classified by gpt-4.1-mini as SGR while the ground truth is TRN. This incorrect classification is shaded orange.

Figure 2. Evaluation results table of two models

segment	code1	code2	code3	sub_seq	video	gpt-4.1-mini	gpt-5-mini
000	TRN	-	-	0		primary_category: TRN secondary_category: OTH confidence: 95 reason: The frames show children entering or preparing in the classroom with teachers assisting, indicative of a transition period. No clear instructional grouping such as whole group or small group activity is taking place. The scene matches the description of within classroom transitions (INS). No other activity type is observed concurrently.	primary_category: TRN secondary_category: OTH confidence: 95 reason: Multiple frames show children entering the classroom with backpacks/coats and adults at the door; overlay text reads 'Arrival' and teacher is guiding students—this indicates a transition into the classroom (arrival).
001	TRN	IND	-	0		primary_category: SGR secondary_category: OTH confidence: 95 reason: The images show multiple children seated at tables with two adults interacting with them closely. The adults appear to be providing instruction or guidance to small groups of children at different tables, consistent with a small group instructional setting. There is no indication of large group circle or carpet time, independent work, or meal time activities.	primary_category: IND secondary_category: OTH confidence: 90 reason: Multiple children seated at tables doing individual worksheets/art with crayons while teachers circulate and monitor/assist. Activity is primarily independent table work rather than whole-group, centers, or transitions.

Note: See orange box in Figure 1 for the human annotation that aligns with this output.

Prompt description and iterations

We used a prompt-based approach to instructing the MLLM how to label video segments. The prompt to the MLLM consists of two parts: the instructions to the MLLM (Figure 3) and the classification indices and descriptions for instructional grouping categories (see Appendix A). These can be used with minor modifications across different families of LLMs. The instructions to the MLLM include a description of the input, the classification description, and a description of the desired output. An output JSON schema is provided to the model to ensure that it generates correct JSON output.

Figure 3. Instructions to the MLLM

```
These are frames sampled from a video at 2 second intervals.
Analyze the provided frames and determine which instructional grouping category best
describes the primary activity shown.


INSTRUCTIONAL GROUPING CATEGORIES
```

```
Each category is represented by a 3-letter code in parentheses.
There is also a sub-category within each category also represented by a 3-letter code in
parentheses.
Choose a primary and secondary category that best match the activity and student arrangement
observed in the video.
Output the primary and secondary categories using the 3-letter codes.
If only one activity is observed, select Other (OTH) for the secondary category.

{instructional_coding_categories}

Provide the output in JSON format with the following keys and description. Do not include
any conversational text or preamble.

{
    "primary_category": [The 3-letter primary category code],
    "secondary_category": [The 3-letter secondary category code],
    "confidence": [A number between 0 and 100 where 0 means low confidence and 100 means
high confidence],
    "reason": [A brief description of what you observed that led to this categorization]
}
```

Results

We evaluated two OpenAI GPT models, 4.1 and 5, on our annotated dataset. Full-length classroom recordings were first broken into segments that annotators identified and labeled. When segments were over 10 minutes in length (i.e., a single instructional grouping episode spanned 10 minutes or longer), they were broken into smaller segments for a total of 132 video segments.

Overall

Overall, GPT-5-mini had a higher score than GPT-4.1-mini (Table 2). We used (1) a **Full** classification for instructional grouping that includes all 10 categories and 7 subcategories, and (2) a **Simple** prompt that includes only the 10 categories to evaluate if the MLLM performed better with fewer categories. The "Primary category" column in Table 2 shows the percentage of segments where the system-produced primary category matched any of the human annotations. The model also outputted a secondary category, and the column "Either category" shows the accuracy when **either** the primary or secondary system-produced category matched any of the human annotations.

Table 2. Summary of OpenAI GPT models with full and simple prompts

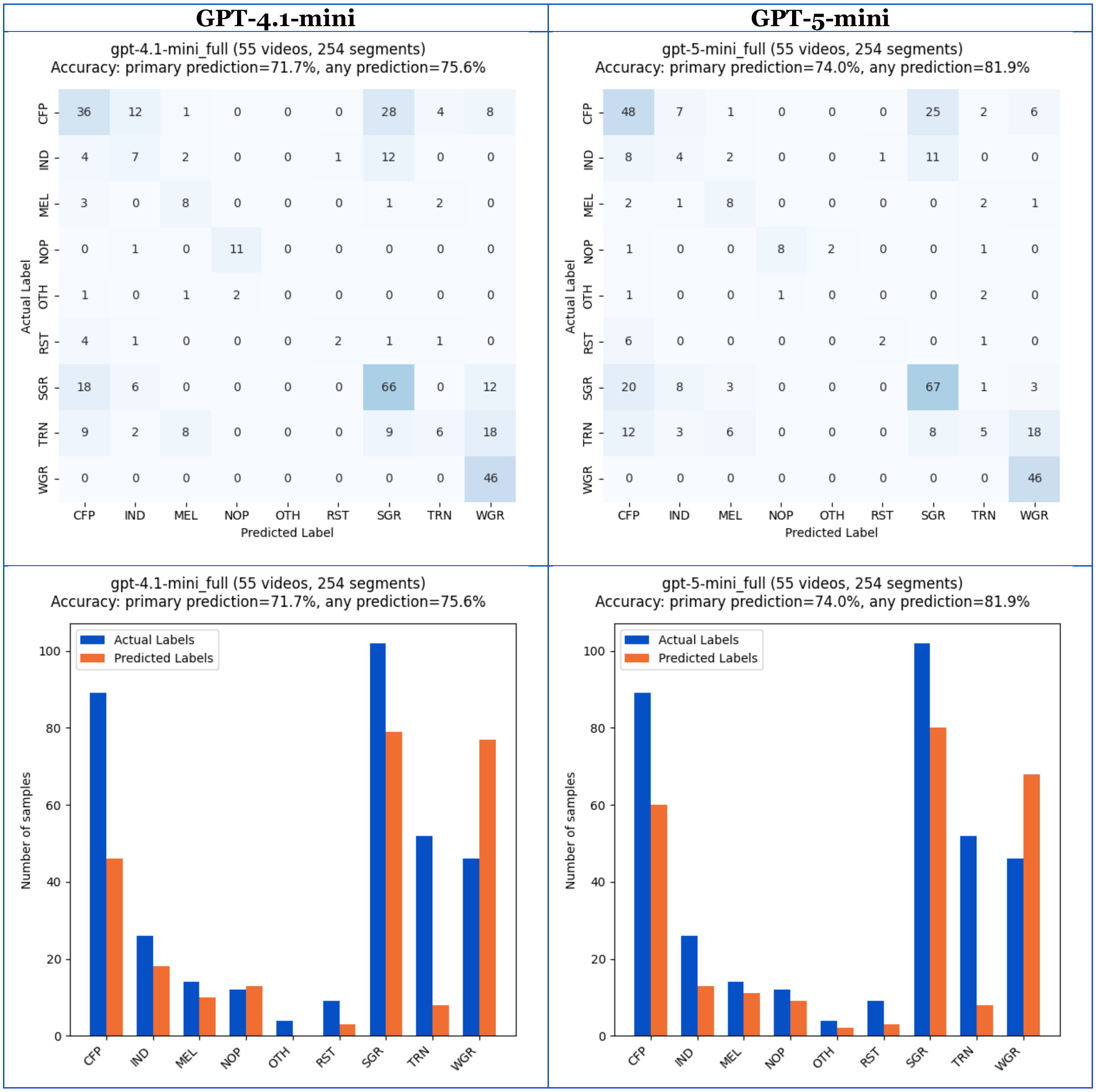
Model	Prompt	Primary category	Either category
GPT-4.1-mini	Full	71.7%	75.6%
GPT-4.1-mini	Simple	56.7%	62.6%
GPT-5-mini	Full	74.0%	81.9%
GPT-5-mini	Simple	75.6%	85.4%

Class-level results

Figures 4 and 5 show more detailed analyses of the evaluation data summarized in Table 2. The figures present confusion matrices that highlight patterns in the classification errors. The "actual labels" (represented by the y-axis) indicate the manually produced ground-truth labels, and the "predicted labels" (represented by the x-axis) indicate the system-produced predicted classes. In the ideal scenario where all predictions are correct, only the diagonal from upper left to bottom right will have nonzero values. For instance, in Figure 4, gpt-5-mini confusion matrices show that SGR (Small Group) labels were correctly classified in 67 out of the total 102 occurrences; SGR was often confused for CFP (Centers/Free Play; 20), IND (Independent; 8) and WGR (Whole Group; 3). The histograms in the second row of the figure indicate the number of occurrences of each class (in blue) and the number of times it was classified correctly (in orange).

A noticeable trend is that all the models consistently underestimated the CFP class while overestimating the WGR class. SGR was commonly misclassified as CFP and IND; ECE experts on the team noted that this misclassification may be due to similar teacher/student makeup in the videos and the descriptive rubric. During SGR, students are seated or arrayed near at least one teacher delivering instruction. When students are similarly arrayed near a teacher but not receiving direct instruction, this class is annotated as CFP. The IND class is similar in that teachers deliver instruction to students but are not required to be near them throughout the period, although they may be. Another misclassification trend is that TRN (Transition) was commonly confused for CFP. Both TRN and CFP contain similar elements of movement, with both teachers and students moving separately or together in the classroom. Misclassification tends to occur when there are similar visual fields of the grouping, but additional context (e.g., transition cue, teacher instruction) may not have been considered by the model.

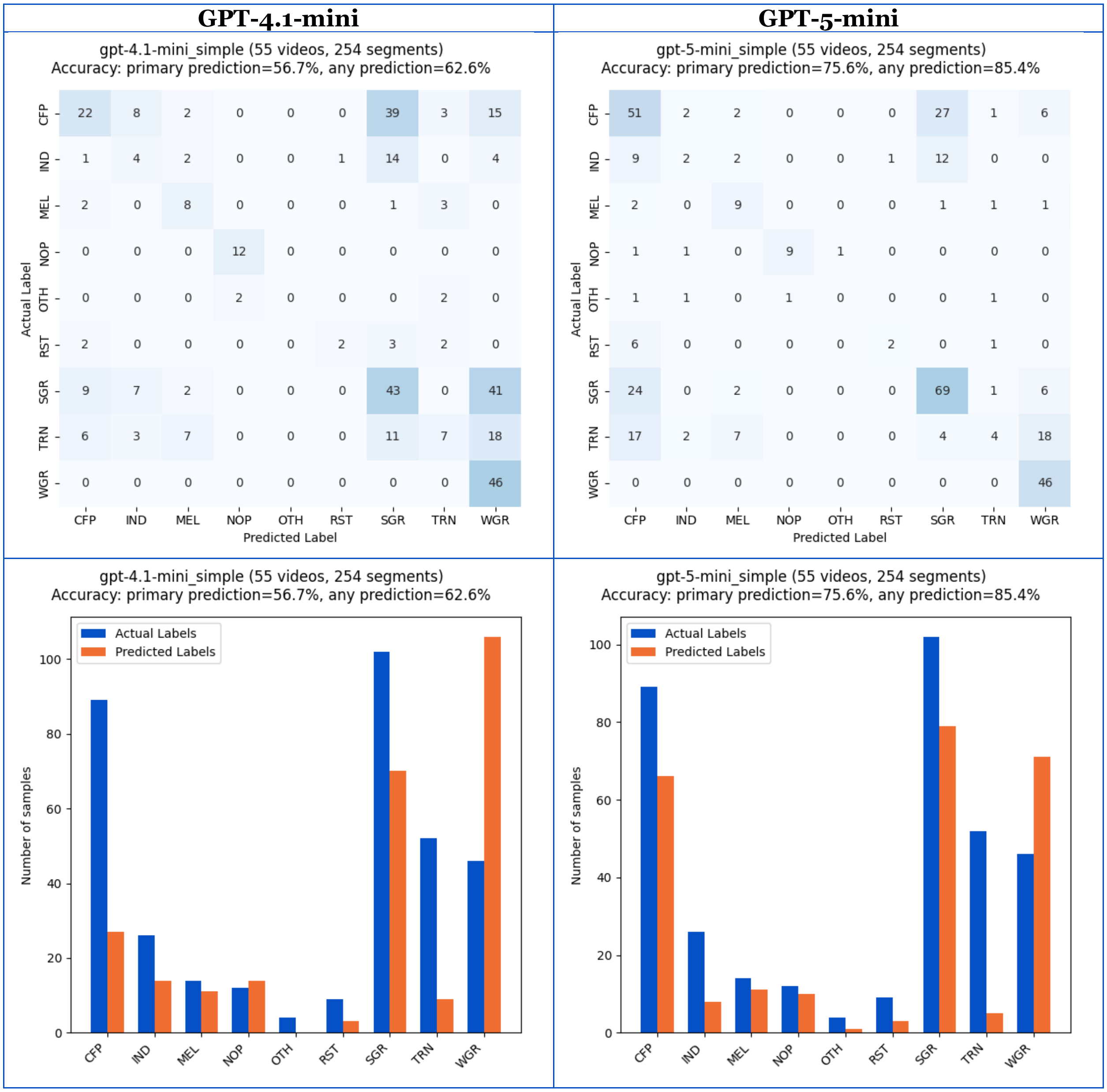
Figure 4. Confusion matrices and class occurrence histograms: Full framework



Simple framework

MLLMs and LLMs sometimes perform better when the prompt is simpler [11]. Therefore, we devised a set of instructions for the instructional grouping classification codes with reduced descriptions and no subcategories. Confusion matrices and class occurrence histograms for the simple models are in Figure 5. We refer to this as the simple framework and compared the MLLMs' performance on the original full framework to this simple framework. Overall, the GPT-5-mini model using the simple framework prompt had the highest accuracy of the four model and prompt combinations, with an accuracy of 85.4% in identifying either category.

Figure 5. Confusion matrices and class occurrence histograms: Simple framework



Limitations

The research detailed in this white paper has limitations to its generalizability. The data used to test the model was small, including 20 classrooms, which does not fully represent the diversity of classroom settings or teacher variability. Additionally, the time represented in dataset is relatively small (20 hours). We offer the approach, framework, and prompt language to encourage further testing and refinement with a larger variety of classrooms and teachers.

Additionally, the goal of the project was to test the feasibility of video as a data source, not to identify the most efficacious prompt and MLLM combination. Therefore, while we developed a prompt with minimal iterations, we did not deeply examine mechanisms to improve the result.

Potential future improvements include exploring additional MLLMs and continuing to revise the language of the prompt and framework with iterative approaches and more diverse data.

The dataset did not have an equal number of observations of each class (e.g., there were more instances of Whole Group than there were of Meal Time). This is only a minor limitation as the MLLM was not trained on the data, so there would be no bias related to the frequency of class observation in the data. In future work, the research team intends to select additional instances of infrequently observed classes to better understand the performance of the MLLM.

Finally, this approach currently does not segment the video sequence automatically, as input is obtained by splitting original video based on timestamps provided by human annotators. The identification of the time or point of transition is important future work for use of this type of approach at scale.

Implications for Future Use

We selected instructional grouping as a classification task because of the significant applications of this model for a variety of organizations and roles within the early childhood ecosystem. The ability to classify the types of instructional grouping being used can be applied to a variety of tasks across the ecosystem. In a local application, teachers, coaches, and center administrators could use this model to identify key moments for review and reflection, understand the dynamics of a classroom (i.e., how much time is spent in different configurations), and support curriculum selection to match the individual culture of teachers and center norms. Additionally, insights from instructional grouping classifications can be useful at scale. For a curriculum provider whose materials are meant to be used in small-group settings, the ability to determine if educators are implementing the materials with fidelity at scale can support alignment with evidence-based approaches and identify opportunities for targeted professional development and feedback, which in turn could improve teacher and student outcomes.

Further, this approach can be used in coordination with another model that SRI developed to identify math and literacy instructional content in early childhood education classrooms [12]. Early childhood is a key piece in the educational puzzle that supports child development. Identifying how, where, and if students are receiving high-quality instruction can open doors for researchers, administrators, and curriculum providers to iterate on materials, build new solutions, and ultimately improve student learning.

References

[1] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv.org. Accessed: Jun. 28, 2024. [Online]. Available: https://arxiv.org/abs/1804.02767v1

[2] “Ultralytics YOLOv8.” Accessed: Jun. 19, 2024. [Online]. Available: https://docs.ultralytics.com/

[3] Y. Tang *et al.*, “Video Understanding with Large Language Models: A Survey,” *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2025, doi: 10.1109/TCSVT.2025.3566695.

[4] S. Curenton, *Conversation Compass: A Teacher’s Guide to High-Quality Language Learning in Young Children*. Databrary, 2018. [Online]. Available: https://databrary.org/volume/732

[5] G. Team *et al.*, “Gemini: A Family of Highly Capable Multimodal Models.” 2025. [Online]. Available: https://arxiv.org/abs/2312.11805

[6] OpenAI *et al.*, “GPT-4 Technical Report.” 2024. [Online]. Available: https://arxiv.org/abs/2303.08774

[7] D. C. Farran and K. Anthony, *Child Observation in Preschools (COP)*. Nashville, TN, USA: Peabody Research Institute; Vanderbilt University. 2014.

[8] C. Bilbrey *et al.*, *Teacher Observation in Preschools (TOP)*. Nashville, TN, USA: Peabody Research Institute; Vanderbilt University. 2007.

[9] R. C. Pianta and B. K. Hamre, *CLASS 2nd Edition Pre-K–3rd Reference Manual*. Teachstone. 2022.

[10] T. Harms *et al.*, *Early Childhood Environment Rating Scale–Third Edition Manual*. New York, NY, USA: Teachers College Press. 2014.

[11] N. F. Liu *et al.*, “Lost in the Middle: How Language Models Use Long Contexts.” *Transactions of the Association for Computational Linguistics*, 12:157–173. 2024. Accessed Oct. 29, 2025. Available: https://aclanthology.org/2024.tacl-1.9/

[12] A. Sundaresan et al. *Developing an AI Model to Identify Math and Literacy Instruction in Early Childhood Education Classrooms* [Technical report]. SRI. 2025

Appendix A. Instructional Grouping Classification

Full instructional grouping classification (includes subcategories)

Figure A1. Instructional grouping classification indices and descriptions

```
1. WHOLE GROUP (WGR)

1.1. CIRCLE TIME (CIR)
Definition: All or most students are standing or seated in a circle or semi-circle with
teacher. Instructional goal not needed.

1.2. CARPET TIME (CAR)
Definition: All or most students are standing or seated in rows on carpet squares with
teacher. Instructional goal not needed.

2. SMALL GROUP (SGR)

2.1. SMALL GROUP (SGR)
Definition: Students are in different areas in classroom (seated, standing) with at least 1
teacher near each group.
Groups of students engaged in either the same or different activity. Teacher has an
instructional goal for the activity.

3. CENTERS AND FREE PLAY (CFP)

3.1. CENTERS AND FREE PLAY (CFP)
Definition: Students are in different areas around classroom (seated, standing, lying).
Students are engaged in activities (e.g., library, water table, dramatic play, blocks) with
or without other students.
There is a maximum of 4 students in each group. A teacher is monitoring. Activities do not
need to have instructional goals.

4. INDEPENDENT  (IND)

4.1. ONE-ON-ONE (UNO)
Definition: Teacher working with individual student for ~30 seconds or more on task or
conversation that is independent from activity that the rest of the class is engaged in. Can
be academic task or supporting child with behavior challenge or emotions.

4.2. INDEPENDENT WORK AT TABLES (TAB)
Definition: Students seated at tables, teacher monitoring. All students engaged in an
individual activity in parallel (e.g., practicing writing names).

5. MEAL TIME (MEL)

5.1. IN-CLASSROOM MEAL TIME (MEL)
Definition: In classroom breakfast, lunch, or snack. Students at tables around classroom.

6. TRANSITIONS (TRN)

6.1. WITHIN CLASSROOM TRANSITIONS (INS)
Definition: Students moving from one instructional grouping to another (e.g., centers to
carpet; carpet to tables; tables to nap cots).

6.2. TRANSITION TO OUTSIDE OF CLASSROOM (OUT)
Definition: Students preparing or lining up to leave classroom (e.g., for recess, end of
day, or specials, like art or PE). Students may have coats or backpacks.

6.3. TRANSITION TO INSIDE OF CLASSROOM (INT)
```

```
Definition: Students entering classroom for morning arrival, or after recess/special.

7. REST TIME (RST)
Definition: Students are lying or resting in different areas in the classroom. The teacher
is monitoring. The classroom volume is low to none, typically 20-90 minutes.

8. OUTSIDE RECREATION (REC)
Definition: Students are outside (e.g., on playground).

9. NO PEOPLE VISIBLE (NOP)
Definition: No people are visible in the video for most of the frames in the video.

10. OTHER (OTH)
Definition: Select this category if the video does not match any of the above categories or
use this for the secondary category if only one primary activity is present.
```

Simple instructional grouping classification

Figure A2. Simple instructional grouping

```
1. WHOLE GROUP (WGR)

Definition: All or most students are standing or seated in a circle or in rows on carpet
with teacher. Instructional goal not needed.

2. SMALL GROUP (SGR)

Definition: Students are in different areas in classroom (seated, standing) with at least 1
teacher near each group. Groups of students engaged in either the same or different
activity. Teacher has an instructional goal for the activity.

3. CENTERS AND FREE PLAY (CFP)

Definition: Students are in different areas around classroom engaged in activities (e.g.,
library, water table, dramatic play, blocks) with or without other students. There is a
maximum of 4 students in each group. A teacher is monitoring. Activities do not need to have
instructional goals.

4. INDEPENDENT  (IND)

Definition: Students are engaged in an individual activity in parallel (e.g., practicing
writing names) or a teacher is working with individual student for ~30 seconds or more
independent from the rest of the class (academic or supporting child with behavior or
emotions).

5. MEAL TIME (MEL)

Definition: In classroom breakfast, lunch, or snack. Students at tables around classroom.

6. TRANSITIONS (TRN)

Definition: Students moving within classroom from one instructional grouping to another
(e.g., centers to carpet). Students preparing or lining up to leave classroom (may have
coats or backpacks). Students entering classroom.

7. REST TIME (RST)
Definition: Students are lying or resting in different areas in the classroom. The classroom
volume is low to none.

8. OUTSIDE RECREATION (REC)
Definition: Students are outside (e.g., on playground).

9. NO PEOPLE VISIBLE (NOP)
Definition: No people are visible in the video for most of the frames in the video.

10. OTHER (OTH)
Definition: Select this category if the video does not match any of the above categories or
use this for the secondary category if only one primary activity is present.
```

Appendix B. Codebook

Figure B1. Instructional Grouping Codebook

Code	Instructional grouping	Definition	Start code when…	End code when…
wgr	**Whole Group**			
cir	Circle Time	Students standing or seated in a ***circle*** or ***semi-circle*** with teacher. Instructional goal not needed.	Teacher begins instructions	Teacher ends instruction; teacher announces transition
car	Carpet Time	Students standing or seated in **rows** on carpet squares with teacher. Instructional goal not needed.	Teacher begins instructions	Teacher ends instruction; teacher announces transition
sgr	**Small Group**	Students in different areas in classroom (seated, standing) with at least 1 teacher near each group. Groups of students engaged in either the same or different activity. Teacher has an instructional goal for the activity.	First student begins activity or teacher begins instructions	Last student stops using materials; teacher announces transition
cfp	**Centers/ Free Play**	Students in different areas around classroom (seated, standing, lying) engaged in activities (e.g., library, water table, dramatic play, blocks) with or without other students (max 4 students per group). Teacher monitoring. Activities do not need to have instructional goals.	First student begins using materials at a center	Last student stops using materials; teacher announces transition
ind	**Independent**			
uno	1 on 1	Teacher working with individual student for ~**30 seconds** or more on task or conversation **independent from what the rest of class is doing.** Can be academic or supporting child with behavior or emotions.	1 student begins activity	Last student stops activity
tab	Independent work (at tables)	Students seated at tables, teacher monitoring. All students engaged in an individual activity in parallel (e.g., practicing writing names).	1 student begins the activity	Last student stops activity
mel	**Meal Time**	In-classroom breakfast, lunch, or snack. Students at tables around classroom.	First student sits at table with food	Last student packs up food
trn	**Transition**			
ins	Within classroom transition	Students moving from one instructional grouping to another (e.g., centers to carpet; carpet to tables; tables to nap cots). Typically 2–5 mins.	Teacher announces transition or visible/audible transition timer begins	Last student completes transition

Code	Instructional grouping	Definition	Start code when…	End code when…
out	Transition to outside of classroom	Students preparing or lining up to leave classroom (e.g., for recess, end of day, or specials, like art or PE). Students may have coats or backpacks.	Teacher calls first student, first student lines up or has materials prepared, teacher announces transition	Last student leaves the room
int	Transition to inside of classroom	Students entering classroom for morning arrival, or after recess/special.	First student enters room	Last student enters room or last parent leaves room (at drop-off).
rst	**Rest Time**	Students lying down in different areas in classroom, teacher monitoring, classroom volume is low to none, typically 20–90 mins.	First child lays on mat	Last child is no longer on mat
rec	**Outside / Recess**	Students are outside (e.g., on playground).	First student is outside	Last student leaves outside
nop	**No People Visible**	No people visible (e.g., if all students and teachers are outside or camera facing a wall for 5+ seconds).		
oth	**Other**	Anything that does not fit into above categories.		