

Automated Discovery of Latent Activity Contexts from Multimodal Sensing for Peer-Interaction Analysis in Preschoolers with Hearing Loss

Shreya Sebastian

Student number: 5957451

MSc Computer Science

at the Delft University of Technology

Thesis committee: Hayley Hung, TU Delft, chair and daily supervisor
Stephanie Tan, TU Delft, daily supervisor
Daniel Messinger, University of Miami
Lynn Perry, University of Miami

Faculty: Electrical Engineering, Mathematics and Computer Science (EEMCS)

This thesis is available at <http://repository.tudelft.nl/>.

Abstract

In inclusive classrooms, the physical presence of children with hearing loss (HL) does not guarantee social inclusion with their typically hearing (TH) peers. Self-report is biased and manual observation is too labour-intensive to cover a full school day at fine temporal resolution. This study adapts a multimodal sensing pipeline that combines ultra-wideband (UWB) spatial tracking with Language Environment Analysis (LENA) audio recordings, applied in a single inclusive preschool with 13 children (6 HL, 7 TH).

A Gaussian Mixture Model fit on four room-level features (adult word count, auditory overlap, displacement, and teacher distance) and selected by a stability-aware rule (lowest mean BIC among K values whose cluster assignments reproduce across random initialisations, mean pairwise Adjusted Rand Index ≥ 0.80) over 2 to 20 components recovers six latent activity contexts, each with a distinct sensor profile. Each context is then given a post-hoc descriptive label drawn from activity types familiar in inclusive preschool classrooms: dispersed transition, peer-driven activity, independent / parallel work, adult-scaffolded peer activity, seated guided work, and whole-class instruction / read-aloud. These labels describe the recovered clusters and are not validated against an external ground truth. Within each context, a linear mixed model with a per-child random intercept compares HL and TH children on three sensor-derivable behavioural markers of inclusion: *peer co-presence* (time in spatial groups, where a *spatial group* is a set of children simultaneously co-present and mutually body-oriented; §5.2.3), *vocal participation rate* (utterances per minute), and *peer affiliation patterns* (time in same-diagnosis vs. mixed-diagnosis spatial groups).

The HL/TH signal distributes differently across contexts for each marker. Peer co-presence shows no substantial HL/TH difference in three of six contexts. HL children exceed TH children in independent / parallel work (+7.8% of the minute, $q < 0.0001$, $d = +1.59$) and fall below TH children in peer-driven activity (−4.3% of the minute, $q = 0.02$) and seated guided work (−5.3% of the minute, $q = 0.002$). Vocal participation rate shows an HL deficit concentrated in the contexts that combine high peer communicative demand with reduced adult scaffolding (peer-driven activity $q = 0.03$ and adult-scaffolded peer activity $q = 0.05$, both significant after false-discovery-rate correction; dispersed transition in the same direction but only marginal, $q = 0.07$), but not in the three contexts where high adult word count or very low auditory overlap slows the pace of exchange. Peer affiliation is the most consistent asymmetry: TH children concentrate grouped time in TH-only spatial groups more than HL children concentrate theirs in HL-only spatial groups in five of six contexts (significantly in four), and HL children spend more grouped time in mixed spatial groups than TH children do across all six (significantly in four). The 6:7 cohort composition produces a baseline TH-above-HL gap of approximately 0.08 on the homophily index under random affiliation; the four significant clusters exceed this baseline (observed gap 0.12–0.19, of which 0.04–0.11 is preference beyond availability), while the two non-significant clusters are within the magnitude expected from cohort composition alone.

Contents

Abstract	i
Nomenclature	iv
1 Introduction	1
1.1 Motivation	1
1.2 Research Questions	2
1.3 Research Contributions	2
2 Background	4
2.1 Vocal Cues and Social Interaction	4
2.2 Peer Interaction in Inclusive Classroom Settings	4
2.3 The Pragmatic Language Barrier	5
2.4 Modality Mismatch and Communicative Preference	5
2.5 Environmental Moderation of Communicative Demand	6
3 Related Work	8
3.1 Automated Measurement of Social Interaction from Sensors	8
3.2 Spatial Group Detection and the Dominant Sets Framework	8
3.3 Prior Work	9
4 Data	10
4.1 Sensor Modalities	10
4.2 Participants and Coverage	11
5 Methodology	12
5.1 Overview and Research Framing	12
5.2 Spatial Feature Extraction	12
5.2.1 Coordinate Pre-processing	13
5.2.2 Per-Frame Displacement	13
5.2.3 Spatial Group Detection via Dominant Sets	13
5.2.4 Aggregation to One-Minute Epochs	15
5.3 Acoustic Feature Extraction	15
5.3.1 LENA Segment Stream	15
5.3.2 Proportional Binning to One-Minute Epochs	15
5.4 Feature Fusion and Temporal Alignment	16
5.5 Discovering Latent Activity Contexts	16
5.5.1 From Per-Child Behavior to a Room-Level View	16
5.5.2 Yeo-Johnson Transformation	17
5.5.3 Gaussian Mixture Model	18
5.5.4 Choice of the Number of Components	19

5.6	Operationalising Social Inclusion per Child, per Context	21
5.6.1	Context-Conditional Comparison Between HL and TH Children	21
5.7	Statistical Inference	21
5.7.1	Model	21
5.7.2	Per-Cluster Tests	22
5.7.3	Homophily Index	23
5.7.4	Exposure Check	23
6	Results	24
6.1	Discovered Latent Activity Contexts	24
6.2	Differential Exposure Across Contexts	27
6.3	Behavioural Markers by Context and Hearing Status	28
6.4	Linear Mixed Model Findings	30
6.4.1	Peer Co-presence	30
6.4.2	Vocal Participation Rate	31
6.4.3	Peer Affiliation Patterns	31
7	Discussion	34
7.1	Findings by Research Question	34
7.2	Methodological Implications	37
7.3	Limitations	37
8	Societal Impact	40
9	Conclusion	41
10	Future Work	43
A	Software and Reproducibility	44
A.1	Software	44
A.2	Reproducibility	44
A.3	AI-Tool Disclosure	44
	References	46

Nomenclature

Acronyms

ASD	Autism Spectrum Disorder
AWC	Adult Word Count
BIC	Bayesian Information Criterion
DHH	Deaf or Hard-of-Hearing
EM	Expectation–Maximization
GMM	Gaussian Mixture Model
HL	Hearing Loss
LENA	Language Environment Analysis
LMM	Linear Mixed Model
TH	Typically Hearing
UWB	Ultra-Wideband

Introduction

1.1. Motivation

Children with hearing loss (HL) often face what this study terms a *pragmatic language barrier*: moment-to-moment breakdowns in turn-taking and emotional-prosody interpretation that can compound into social neglect by typically hearing (TH) peers [1]. The underlying pragmatic difficulties are well documented [2, 3].

Peer interactions between HL and TH children are shaped by a *modality mismatch*, a dynamic in which the dominant communicative channel, most often rapid verbal exchange, is not equally accessible to all participants. The verbal dominance of more communicatively capable peers can unintentionally exclude children who rely more on embodied or non-verbal strategies during play [4] (Ferreira’s observations were of children with intellectual disabilities; the dynamic is taken up in the HL/TH setting in §2.4), and the resulting communication gaps appear as lower social acceptance and a higher risk of loneliness [5]. The mechanisms of isolation are often subtle: a small failure of response contingency, an ignored bid for interaction, rather than open rejection [6, 3].

Traditional measurement approaches, self-reports and manual observation, struggle to capture the fine-grained, continuous interactions that shape this experience [7]. What is needed instead are objective tools that can identify the activity contexts in which these barriers are most pronounced [8]. This study addresses that need by adapting an automated, multimodal sensing and analysis pipeline to investigate social inclusion of HL children in naturalistic inclusive classrooms, building on earlier work from the same research group [9, 10]. The relationship to that earlier work is described in Chapter 3.

Children with hearing loss in mainstream classrooms are substantially more likely to be socially isolated than their typically hearing peers [1, 11, 5], and broader work on social-communication difficulties indicates similar patterns for children with related pragmatic-language profiles [6]. The core problem addressed here is that social inclusion is not a stable per-child trait but varies with the prevailing classroom activity. A daily average over context-specific effects of opposing sign can obscure both, which may account for the limited signal detected by pooled HL/TH comparisons and for the difficulty of localising the situations in which exclusion concentrates. Identifying those situations, and measuring inclusion within them rather than across the whole day, is the open problem this study targets. Wearable sensing addresses the limits of manual methods by recording continuous streams of

physical co-presence and vocal activity. The computational challenge is to convert these streams into meaningful units of analysis: activity contexts within which inclusion can be compared across children.

1.2. Research Questions

This study asks two connected questions, each driving specific design and analysis choices in the sections that follow.

RQ1.

Can a set of interpretable latent activity contexts in an inclusive classroom be recovered from fused ultra-wideband (UWB) spatial data and Language Environment Analysis (LENA) acoustic data using unsupervised clustering?

RQ2.

Within each recovered activity context, do children with hearing loss and their typically hearing peers differ on three sensor-derivable behavioural markers of inclusion: (a) peer co-presence, (b) vocal participation rate, and (c) peer affiliation patterns?

The following three hypotheses are *a priori* predictions derived from the modality-mismatch literature reviewed in Chapter 2, not findings of this study; they are tested in Chapter 6. The literature's central prediction is that HL children compensate for reduced access to rapid verbal exchange by staying physically close to peers while speaking less [12, 4, 13]. Each hypothesis is stated below with its rationale.

- H1.** *Peer co-presence will not show a consistent HL deficit across contexts; where contexts produce HL/TH differences, HL children will sustain or exceed TH co-presence in contexts of relaxed communicative demand, operationalised here as low auditory overlap or activity in which adult speech mediates peer turn-taking. Rationale: HL children sustain physical co-presence with peers as a compensatory strategy rather than withdrawing from peer groups [12, 4].*
- H2.** *The HL/TH gap in vocal participation is context-dependent. Rationale: the gap should be largest where fast, multiparty peer talk dominates with little adult mediation of turn-taking, and smaller where a high volume of adult speech or low background overlap slows the exchange [13, 8, 2, 3].*
- H3.** *TH children affiliate more with same-status peers than HL children do, so HL children spend relatively more of their grouped time in mixed-diagnosis groups; and this asymmetry is larger in contexts in which children choose their own groupings (e.g. peer-driven activity) than in teacher-organised ones (e.g. whole-class instruction, seated guided work). Rationale: Tian et al. [14] found same-status affiliation preferences at the group level in an inclusive HL/TH preschool, and homophily is generally stronger where group composition is self-selected rather than externally imposed [15] (McPherson's review covers adult social networks, and the prediction here extrapolates that pattern to preschool peer formation).*

1.3. Research Contributions

This study makes two contributions.

1. **Unsupervised discovery of activity contexts whose sensor profiles are descriptively consistent with phases of the classroom day.** A Gaussian Mixture Model (GMM) over four room-level features, with the number of components K selected by a stability-aware BIC rule over $K = 2\text{--}20$

(§5.5.4), recovers six latent activity contexts with distinct sensor profiles. Each is given a post-hoc descriptive label corresponding to a recognisable phase of the school day (§6.1). Whether the six-context decomposition generalises is a question for multi-site replication (§7.3).

2. **Empirical evidence that HL/TH differences in these markers are context-specific rather than uniform.** Each of the three markers (peer co-presence, vocal participation rate, and peer affiliation patterns) is compared between HL and TH children separately within each discovered context, using a linear mixed model (LMM) that first removes stable differences between individual children. The HL/TH difference is not constant across the day: for each marker it is statistically significant in some contexts and negligible in others, so whether HL and TH children differ depends on the prevailing classroom activity. The full per-context results, including the statistical tests, are reported in Chapter 6 and summarised in the Conclusion.

2

Background

This chapter introduces the developmental and communicative phenomena that motivate the measurement choices made later in this thesis. Three ideas run through it: peer interaction is organised around a dominant communicative channel; that channel is not equally accessible to all children; and which channel a child can or prefers to use is itself shaped by the surrounding activity.

2.1. Vocal Cues and Social Interaction

Peer interaction is a primary developmental context in early childhood. A meta-analysis of 71 studies reports a small-to-moderate gain in learning outcomes from peer collaboration over non-collaborative conditions (Hedges' $g = 0.40$) [16]. The effectiveness of these interactions depends on the communication channels through which social meaning is negotiated, with vocalisation a central component. Beyond literal meaning, communication carries paralinguistic features such as pitch, intonation, speech rate, and loudness [17, 18], which listeners use to infer a speaker's emotional and cognitive states and to judge intentions, attitudes, and sincerity [17]. These acoustic properties reliably signal emotional states (anger, sadness, fear, happiness) and stable speaker characteristics (sex, age, perceived dominance) [19, 20]. Speakers actively adjust vocal features to communicate intentions. For example, they modulate pitch to signal dominance or submissiveness [20].

A specific barrier for HL listeners is *emotional prosody*, the vocal patterns that carry emotion and intention. Because emotional prosody is what allows listeners to distinguish sarcasm from sincerity [21], difficulty processing these cues is an important barrier to successful social communication [3]. Cochlear-implant users illustrate this: they perform above chance at vocal emotion recognition but worse than typically hearing listeners, because the limited spectral and temporal resolution of cochlear implants restricts access to key acoustic markers such as pitch variation, voice quality, and articulation patterns [22]. Children who are better at interpreting and using vocal and non-verbal pragmatic signals are seen as more socially competent and are more often accepted as play partners [23].

2.2. Peer Interaction in Inclusive Classroom Settings

Inclusive education places students of differing abilities in the same classroom [24]. The resulting peer-interaction dynamics differ from those of homogeneous settings, particularly where children with

communication-affecting developmental delays are educated alongside typically developing peers.

Qualitative studies and reviews find that HL students often have trouble entering the fast-paced group conversations common among TH peers. This can lead HL children to prefer non-verbal play, such as physical play, or to seek out solitary activities, behaviours that TH peers can misread as a lack of interest in interaction, creating a barrier between the two groups [12]. Even when HL children have age-appropriate language skills, their attempts to interact through eye gaze or turn-taking are often ignored by hearing peers, and HL children are consequently more likely to be neglected and to become isolated [1].

2.3. The Pragmatic Language Barrier

Pragmatic language, the social use of communication, is a major challenge for many HL children. Compared to their TH peers, HL children often show delays in pragmatic language skills even when other cognitive and language functions are within the average range [3, 25]. These delays appear as difficulty managing turn-taking, inferring a speaker's intention, responding contingently, and interpreting emotional prosody. Such difficulties are strongly linked to higher rates of social isolation and peer rejection [2, 3]. Following Antia et al. [2] and Mood et al. [3], the term pragmatic language barrier is used here as shorthand for the cumulative effect of these difficulties on peer access.

Several mechanisms show how this barrier manifests in peer interactions. First, vocalisation patterns shape peer engagement. Fasano et al. showed that children with autism spectrum disorder (ASD) occupy more peripheral positions in classroom social networks (lower degree centrality) and produce fewer peer-directed vocalisations than typically developing or developmentally delayed peers, and that classroom vocalisation centrality is associated with end-of-year language outcomes. Dyadic exchanges are also reciprocal, with children vocalising more to peers who had vocalised more to them in prior observations [26]. Second, breakdowns often occur through failures of response contingency. Murphy et al. found that children identified as having low pragmatic skills were more likely to give ignoring responses to peer questions [6], and Bobzien et al. reported that HL children produced fewer initiated verbal comments than their hearing peers (group means 1.84 vs. 7.83; per-pair test statistics in the original paper) but were generally able to maintain conversational turns once interactions had started [13]. Third, pragmatic language acts as a key mediator of inclusive peer interaction. Lin et al. showed that the likelihood of cross-status play is strongly predicted by pragmatic language abilities, and that when pragmatic language and self-regulation are statistically controlled, the effect of disability status on interaction frequency disappears [27]. Communication style, on this view, matters more for cross-status play than diagnostic category alone.

2.4. Modality Mismatch and Communicative Preference

The pragmatic-language literature describes what HL children find hard in peer interaction. A related body of work, central to this study, describes what happens when difficulty in one channel pushes interaction toward another. *Modality mismatch* is the term used here for the phenomenon Da Silva et al. [12], Ferreira [4], and Bobzien et al. [13] describe: a dynamic in which the dominant communicative channel, most often rapid spoken exchange among TH peers, is not equally accessible to all participants. *Modality preference* is a child's matching tendency to seek out or avoid a given channel. The two are paired: on the framing adopted here, a mismatch at the group level is partly a function of which channel each child can use, and apparent preference is partly a product of which channels the other children present leave open.

Three lines of evidence support this framing for HL/TH peer interaction.

- **Loss of multiparty talk.** HL children often face difficulties entering and following the multiparty exchanges that hearing peers use as a key source of social information [1].
- **Verbal vs. embodied interaction.** Ferreira observed dyads of typically developing children and children with intellectual disabilities (not hearing loss) and reported that verbalization by typically developing children can block collaboration when used to reprimand or control, shutting out the participation of children with communicative disabilities. The same dyads collaborated more smoothly under *embodied collaboration*, a non-verbal mode in which the shared desire to play took priority over verbal task completion [4]. The population differs from the HL/TH case here, but the underlying dynamic is the same one this study examines: a verbal channel that is not equally usable, alongside a non-verbal mode that re-opens collaboration.
- **Environmentally induced mismatch.** The physical environment can itself create mismatch by demanding specific vocal behaviours. Irvin et al. found that the rate of vocal interaction varied with the type of activity area children were in [8]: some contexts favour vocal exchange, whereas others permit parallel or non-verbal participation.

Central to this thesis is the idea that modality preference is not a stable trait of a child but a joint function of the child and the surrounding activity. Da Silva et al. document HL children moving toward physical play during playground free time. In that setting the verbal channel is plausibly closed by noise and peer conversation speed, so what appears to be a stable preference for non-verbal play may be partly a response to a communicative environment in which the verbal channel is less viable [12]. Bobzien et al.'s observational study of dyadic play between HL and TH preschoolers [13] shows the converse: even in a single quiet, table-based setting, HL children produce far fewer initiated verbal comments while still maintaining verbal and play turns once an exchange is underway. The implication is that one should not expect a single uniform HL/TH gap in vocal or physical participation, but rather a set of gaps whose size and sign depend on the communicative demand of the surrounding context. The methodology in Chapter 5 is structured to test this prediction by discovering those contexts from sensor data rather than defining them in advance.

Successful interaction often relies on adult scaffolding [28]. Typically developing peers align to the sensory capacities of children with Down syndrome or hearing loss by adjusting pace and gaze to keep the interaction going [29]. Scaffolding is a special case of the broader point above: a deliberate redesign of communicative demand so that a child's accessible modality becomes the dominant modality of the group.

2.5. Environmental Moderation of Communicative Demand

Communicative demand varies systematically with classroom activity. Four recurring dimensions of this variation are relevant for inclusive classrooms and are operationalised in this thesis as the four room-level features used for context discovery (§5.5). The density of adult speech determines whether interaction is instruction- or peer-dominated. The rate of auditory overlap determines how fast-paced and multiparty peer exchange is, and the magnitude of competing acoustic signal the listener must filter. The amount of physical movement determines whether children are seated in proximity or spread out and mobile. The spatial configuration of children relative to the teacher serves as a proxy for whether an activity is adult-led and teacher-proximal or child-led and teacher-distant. Irvin et al. show

this empirically through activity-level vocalisation differences [8], and Ferreira shows it qualitatively through verbal-control vs. embodied-collaboration phases in the same dyad [4]. This thesis takes the position that the environment does a substantial share of the explanatory work, since who is included at a given moment depends not only on which child is present but on the ongoing classroom activity. Inclusion must therefore be measured within context, and the contexts themselves should be defined by the features the literature identifies as axes of communicative demand.

Beyond peer-interaction dynamics, the acoustic environment itself disadvantages HL listeners asymmetrically. Children with hearing loss require substantially more favourable signal-to-noise ratios than TH peers to achieve equivalent speech recognition, and this gap widens in the reverberant, multi-talker conditions typical of preschool classrooms [30]. From the perspective of communicative demand, classroom minutes with overlapping speech and minutes with high background noise both reduce the intelligibility of speech that an HL child needs to follow peer talk, particularly during peer-driven exchange where adult clarification is sparse. The two are therefore operationalised here as a single axis of acoustic demand, a choice implemented in §5.3 by bundling LENA's overlap and noise segment categories. Whether they are in fact dissociable phenomena under finer-grained analysis is an empirical question not addressed here.

Implications for the methodology. The background sets up three measurement targets. First, contexts vary in communicative demand, so inclusion must be measured within context rather than averaged over the day. Second, physical co-presence in peer groups and vocal participation are distinct behavioural aspects of inclusion that can come apart. Third, modality preference is a joint property of child and activity, so an HL/TH gap can point in different directions in different contexts. A fourth, statistical target, separating individual differences in baseline sociability from group-level differences, is developed in §5.7.

3

Related Work

3.1. Automated Measurement of Social Interaction from Sensors

In child development, Eichengreen et al. [31] showed that wearable proximity sensors capture differences in schoolyard social participation between DHH (deaf or hard-of-hearing) students and hearing peers, and reveal social difficulties invisible to traditional peer-nomination methods. Their setting is outdoor and uses coarse-resolution proximity sensing. This study applies UWB tracking and child-worn LENA audio inside the classroom at 100 ms and 1 min resolution respectively, allowing spatial group membership and vocal participation to be jointly observed. LENA itself [32, 33, 34, 35, 36] has become the standard for continuous child-worn audio capture, providing validated labelling of who is speaking, along with word and utterance counts. Much of the existing work operates at the level of a single outcome (vocalisation rate, proximity, or network centrality) aggregated over the whole observation window. Relatively little has explicitly modelled the context-dependence of communicative demand highlighted in §2.5.

A parallel line of work uses vocalisation and positional data to measure peer engagement in inclusive classrooms. Fasano et al. [26] showed that peer-directed vocalisations from LENA are a fine-grained correlate of inclusion for children with ASD. Irvin et al. [8] combined LENA with indoor positioning to show that the physical zone of the classroom systematically shapes vocal interaction. This study follows the tradition of combining proximity and acoustics, and extends it specifically to an inclusive HL/TH setting.

3.2. Spatial Group Detection and the Dominant Sets Framework

A central computational problem is identifying, from proximity and orientation data, which children are jointly engaged at a given moment. The motivating construct is the *F-formation* [37, 38]: a spatial arrangement in which two or more people orient their bodies to create and sustain a shared interactional space. The F-formation is only one operationalisation of the broader notion of a jointly engaged group, and it has been characterised and validated primarily in free-standing adult conversation [39, 38]; whether young children form true F-formations has not been established. For this reason the objects detected in this study are not claimed to be F-formations. They are referred to as *spatial groups*: sets of children who are simultaneously close together and mutually body-oriented. A spatial group is used as

a non-verbal proxy for joint engagement, motivated by the F-formation idea rather than equated with it.

Group membership is recovered with the Dominant Sets framework [40], which extracts maximally cohesive subsets of a pairwise-affinity graph without fixing the number or size of groups in advance [38]. This study makes no methodological contribution to group detection itself: it reuses the detection approach and implementation of Tian et al. [9, 10] so that the recovered spatial groups are directly comparable to those in the earlier study of this cohort. The detailed formulation is given alongside the rest of the method in §5.2.3. Tian et al. [14] themselves caution that it is not known whether children of this age form true F-formations as observed in adults; this study inherits that caution and treats the detected objects only as spatial groups. The established detector is retained for comparability with this prior work rather than as a claim to state-of-the-art group detection.

3.3. Prior Work

The pipeline in this study builds directly on two earlier outputs of the same research group, referred to together as the *reference pipeline*: an open-source group-detection implementation [10], which supplies the Dominant Sets extractor, and a follow-up study of homophily in inclusive-classroom interactions [9]. One methodological difference between this pipeline and the reference pipeline should be stated at the outset. The published analysis of Tian et al. [9] detects groups from a proximity-only affinity, and does so deliberately: it avoids any assumption about how children orient themselves, on the grounds that young children’s body orientation is an unreliable cue. The open-source implementation this study builds on instead parameterises the affinity as a proximity Gaussian multiplied by a clipped mutual-orientation term, with $\sigma = 0.5$ m (Eq. 5.2), and it is this implementation that the present pipeline uses. Including the orientation term is therefore a deliberate departure from Tian et al.’s published choice; its consequences are discussed in §7.3.

The contributions of this study relative to the reference pipeline are (i) GMM-based unsupervised discovery of latent activity contexts from room-level features, (ii) a per-cluster LMM with a per-child random intercept that separates context-level from child-level variation in the inclusion outcomes, and (iii) the three-marker decomposition of inclusion into peer co-presence, vocal participation rate, and peer affiliation patterns.

Two limitations of the reference pipeline carry into the results reported here. First, the mutual-orientation term is computed from a single heading vector. Detectors of conversational groups typically use a cone of attention [39], whereas the single-vector version used here may be especially unsuitable for young children, whose head and shoulder orientation are noisy and not always aligned with their focus of attention [41]. Second, affinity is computed for pairs and then clustered, so groups of three or more are recovered only indirectly through the Dominant Sets step. Both are revisited in Chapter 7.

4

Data

The analyses use inclusive preschool observations combining UWB indoor positioning, child-worn LENA audio recorders, and a demographic and clinical mapping. The data were collected and de-identified by researchers at the University of Miami before being shared for this analysis; all records use anonymised participant codes and contain no names.

4.1. Sensor Modalities

Ultra-wideband positioning. Each child and teacher wears a vest carrying two Ubisense UWB tags. The system reports each tag’s position in metres relative to a fixed room frame at 10 Hz (100 ms grid). A child’s position (x, y) is taken as the midpoint of their two tags, and the *shoulder-orientation angle* θ (in degrees, relative to the room’s x -axis) is derived from the line joining them. This θ is a shoulder (body) orientation, not head direction or gaze. The position stream was Kalman-smoothed during the original data collection and preprocessing, before the data were shared for this analysis; the present pipeline therefore starts from these Kalman-smoothed 10 Hz positions (§5.2).

LENA acoustic recorder. Each child wears a LENA digital language processor that continuously records audio and, offline, classifies every moment into speaker-type segments [32, 33]. Each recording day produces a per-child log of time-stamped segments carrying the speaker type (including overlap and noise categories), the number and total duration of child vocalisations, and adult word counts. LENA has been validated against human transcription for child vocalisations in naturalistic settings [33, 35].

Demographic and attendance records. A static demographic record links each child’s de-identified participant code to a hearing status (HL or TH). A separate per-day attendance record flags each (child, recording day) as present or absent and carries the physical recorder identifier used by that child on that day, since recorders are reassigned across children on different days. Inference is limited to (child, day) pairs flagged present. Teacher versus child status is inferred from a tag-naming convention in the participant code rather than from a demographic field. No names are stored.

Manual activity coding. An independent manual coding of the daily classroom routine is also available. It was produced through live in-classroom observation by the University of Miami research team that collected the data. Each period of the day is assigned to one of eight activity types: personal care (bathroom), meal time, circle time, story time (book reading), therapy, transition, organised playtime, and free play. This coding is used only as a loose external reference point when interpreting the data-driven activity contexts (§5.5.4); it is never used to fit or constrain the clustering model.

4.2. Participants and Coverage

The analyses use recordings from a single inclusive preschool classroom with 13 children (6 HL and 7 TH) together with their teaching staff, in a room whose instrumented area spans roughly 13×9 m. Recordings were collected over 13 days, with a mean of about 164 minutes of recording per child per day (SD 36).

How these raw recordings are turned into per-minute features, fused across the two modalities, and filtered to the minutes used for analysis is described in Chapter 5. The fused spatial–acoustic stream yields 22,730 jointly observed (child, minute) rows. Restricting to minutes in which at least three children are jointly observed (required for the room-level context discovery of §5.5) leaves 22,251 minutes labelled with an activity context, of which 21,847 fall on days the attendance record confirms the child was present; only these are used for the statistical models. Minutes are dropped only when the attendance log records the child as absent or has no record for that day, so the analysis rests on attendance-confirmed minutes rather than on sensor activity alone.

5

Methodology

5.1. Overview and Research Framing

This chapter describes the pipeline that turns the two raw sensor streams into the dataset used for analysis: a table in which each row is one (child, minute) pair, labelled with an activity context, on which HL/TH differences can be tested within context. The two streams are fused because neither alone is enough: the UWB position data are needed to see physical co-presence and grouping, and the LENA audio is needed to measure vocal participation. The pipeline runs in stages, and each stage maps onto a research question.

- **Spatial feature extraction** (§5.2) and **acoustic feature extraction** (§5.3) produce the peer-co-presence and vocal-participation-rate signals needed by RQ2, together with the spatial group membership data from which the peer affiliation metrics are derived.
- **Room-centric feature construction and GMM clustering** (§5.5) operationalise the “latent activity context” of RQ1.
- **LMM inference** (§5.7) fits a per-child random intercept so that between-child variation in each outcome is absorbed before the fixed-effect HL/TH contrasts are estimated.

In detail, the two fused inputs are:

- (i) **UWB spatial positioning**, providing (x, y, θ) triples per participant at 10 Hz.
- (ii) **LENA acoustic segments**, providing per-segment counts of child utterances, adult words, and overlap/noise durations.

Demographic metadata maps each child to a hearing status (HL/TH).

5.2. Spatial Feature Extraction

This section turns the UWB position stream into the per-child spatial quantities used later: *cumulative displacement* (how much a child moves, §5.2.2) and *spatial-group membership* (whether, and with whom, a child is in a co-present, body-oriented group, §5.2.3), together with each child’s distance to the nearest

teacher. Cumulative displacement and teacher distance later feed the room-level features for context discovery (RQ1); spatial-group membership feeds the peer co-presence and peer affiliation markers (RQ2). The first step is to put the raw positions onto a regular time grid.

5.2.1. Coordinate Pre-processing

The data arrive already resampled and Kalman-smoothed as part of the original data collection (Chapter 4), so the UWB stream provides at most one (x, y, θ) sample per child every 100 ms. The only coordinate handling performed here is a light cleanup: where more than one reading falls in the same 100 ms interval for the same child, the readings are de-duplicated to a single sample, and each timestamp is snapped to the nearest point on the 100 ms grid. Snapping moves a timestamp by at most 50 ms (half the grid spacing), which is far below the timescale over which spatial groups form and dissolve (seconds) and so does not affect group detection.

5.2.2. Per-Frame Displacement

To capture how physically active a child is, the Euclidean displacement between consecutive samples is computed as

$$\Delta p_i(t) = \sqrt{(x_i(t) - x_i(t-1))^2 + (y_i(t) - y_i(t-1))^2}, \quad (5.1)$$

on the regular 100 ms grid. This is a distance moved between two samples, not a rate (it is deliberately not called “velocity”, because dividing by the very short time between samples would mainly amplify position noise). Summed over a minute, $\sum_t \Delta p_i(t)$ gives the child’s cumulative displacement for that minute, i.e. the total distance moved: high for a child walking around, low for one who stays put. To guard against spurious displacements across extended gaps (for example when a child leaves the instrumented area and returns several meters away), any displacement computed across a gap longer than 5 s is set to zero. Beyond 5 s, the most common cause of a position jump is the child leaving and re-entering the instrumented area rather than a genuine continuous movement. The asymmetric cost structure motivates erring toward zeroing: a missed movement biases displacement downward by at most a few metres in the affected minute, whereas an unfiltered cross-room jump can inflate that minute’s displacement by tens of metres and shift its cluster assignment.

5.2.3. Spatial Group Detection via Dominant Sets

As defined in Chapter 3, a spatial group is a set of children who are simultaneously close together and mutually body-oriented, used here as a non-verbal proxy for active joint engagement. Spatial groups are detected with the procedure and implementation of the reference pipeline [9, 10], adopted unchanged: a pairwise socio-spatial affinity (below), followed by Dominant Sets extraction [40]. Neither component is novel to this study.

Pairwise socio-spatial affinity. For every 100 ms frame with n children, a pairwise affinity matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is built. Affinity between children i and j is the product of a spatial-proximity term and a mutual-orientation term [9].

$$A_{ij} = \underbrace{\exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right)}_{\text{proximity}} \times \underbrace{\max(0, \hat{u}_{ij} \cdot \hat{d}_i) \max(0, -\hat{u}_{ij} \cdot \hat{d}_j)}_{\text{mutual orientation}}, \quad (5.2)$$

In this expression, $d_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\|_2$, $\sigma = 0.5$ m, $\hat{\mathbf{d}}_i = (\cos \theta_i, \sin \theta_i)$ is the unit orientation vector of child i , and $\hat{\mathbf{u}}_{ij} = (\mathbf{p}_j - \mathbf{p}_i)/d_{ij}$ points from i to j . The mutual-orientation term is zero unless both children’s orientation vectors have a positive projection onto the line between them. The bandwidth $\sigma = 0.5$ m sets the spatial scale on which co-presence is detected: the proximity term falls to ≈ 0.61 at 0.5 m, ≈ 0.14 at 1.0 m, and ≈ 0.01 at 1.5 m, so it effectively recognises children within personal-distance range and rapidly discounts pairs beyond it. This matches the proxemic distances Kendon’s joint-engagement framework identifies (§3.2) [37], scaled to preschoolers, who interact at shorter distances than adults. The value is also the one used in the reference implementation of Tian et al. [10]; retaining it keeps the spatial-group definition directly comparable to that prior study of the same cohort. The product form requires both terms to be satisfied for a non-trivial affinity: two children who are close but back-to-back, or oriented toward one another but several metres apart, both score zero. This matches the requirement reviewed in §3.2 that joint engagement requires simultaneous spatial and orientational alignment [37, 38]; an additive combination would let either term alone produce a high affinity, conflating proximity-only and orientation-only configurations that the construct treats as distinct. The affinity is best read as a proxy for body-pose-consistent joint engagement rather than direct mutual gaze. The published detector of Tian et al. [9] uses proximity alone; the multiplicative orientation term here is an addition specific to the implementation this study reuses. Both the rationale for including it and its consequences for the peer-co-presence measure are revisited in §7.3.

Dominant Sets extraction. Spatial groups are extracted as dominant sets of $\mathbf{A}$ by iteratively solving

$$\max_{\mathbf{x} \in \Delta} f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}, \quad \Delta = \{\mathbf{x} \in \mathbb{R}_{\geq 0}^n : \mathbf{1}^\top \mathbf{x} = 1\}, \quad (5.3)$$

via replicator dynamics [40]. At convergence, participants with support $x_i > \epsilon$ (in the implementation, $\epsilon = 10^{-5}$) are flagged as members of one spatial group. Members are removed from $\mathbf{A}$ and the procedure repeats until no further cohesive subset is found. Three implementation thresholds matter for reproducibility, all inherited from the reference implementation [10] and consistent with the values used in Pavan and Pelillo’s original formulation [40]. (i) The replicator iteration is stopped early if the largest entry of $\mathbf{A}$ is below 0.01; under $\sigma = 0.5$ m this corresponds to no pair of children being closer than ≈ 1.5 m while mutually oriented, well beyond the proxemic range over which spatial groups are defined. (ii) The membership cutoff $\epsilon = 10^{-5}$ is well below any value that would alter group membership: a participant assigned support below 10^{-5} in the unit simplex contributes negligibly to the dominant-set quadratic form. (iii) The replicator iteration starts from the uniform point of the simplex (every participant weighted equally, as in the standard formulation of Pavan and Pelillo) and runs to a fixed tolerance of 10^{-6} under a 1000-iteration cap; in practice convergence is reached in far fewer than 1000 iterations and the cap is a safeguard. Extracted subsets with fewer than two members are discarded, consistent with the spatial-group definition requiring two or more participants [38]. The extractor has no random component, and the spatial-feature table reproduces exactly across runs.

Dominant Sets is preferred to a partitioning method such as k -means because spatial groups in a free-play classroom vary in size, from pairs to larger groups, and need not assign every child in the room to a group.

Group demographic classification. Each detected spatial group is labelled *HL-only*, *TH-only*, or *mixed* from the hearing statuses of its members. This is what makes the affiliation marker of RQ2 measurable: whether HL and TH children tend to group with same-status or with mixed-status peers. HL children

are expected to be more socially included when they spend grouped time in mixed-status groups rather than in HL-only sub-groups [42, 43].

5.2.4. Aggregation to One-Minute Epochs

Per-frame (100 ms) indicators are aggregated to one-minute epochs per child. Cumulative displacement is summed across the minute, mean teacher distance is averaged, and the four binary group-membership flags (in any spatial group, in an HL-only, TH-only, or mixed-composition spatial group) are averaged over the frames in the minute and expressed as percentages. The one-minute resolution is chosen to span multiple LENA conversational turns (whose mean and median durations are 1.66 s and 1.05 s respectively; §5.3) so that the per-minute acoustic features stabilise to activity-level rather than turn-level variation, while remaining short enough that distinct activity phases of the school day do not blur together. Finer windows (e.g. 10 s) would frequently fall inside a single LENA turn, so the variance in the acoustic features at that scale would be dominated by where turn boundaries happen to fall rather than by activity differences. Coarser windows (e.g. 5 min) span multiple distinct activity phases, defeating the purpose of context detection.

5.3. Acoustic Feature Extraction

5.3.1. LENA Segment Stream

LENA classifies every moment of audio into speaker-type segments [32, 33, 34]. Four segment-level quantities are used here.

- **Child utterance count**, the number of discrete child vocalisations within the segment.
- **Child utterance duration**, the total time (seconds) of child vocalisations within the segment.
- **Adult word count (AWC)**, the number of discrete adult words heard by the child within the segment.
- **Auditory overlap**, the total duration of segments classified by LENA as overlapping speech (multiple simultaneous talkers) or background noise (non-speech acoustic interference), in either case at near- or far-field range relative to the recorder (LENA categories OLN/OLF for overlap and NON/NOF for noise) [34]. The four LENA categories are bundled into a single feature on the grounds developed in §2.5: both reduce the intelligibility of speech that an HL listener needs to follow, and they are therefore operationalised here as a single axis of acoustic demand. Whether the two are dissociable under finer-grained analysis is not addressed in this work. The label *auditory overlap* (in preference to *vocal overlap*) reflects this broader scope.

5.3.2. Proportional Binning to One-Minute Epochs

LENA reports its measurements over speaker segments whose length varies, beginning and ending whenever speech does rather than on whole-minute boundaries. These segments are short relative to a one-minute epoch: across the 1,752,443 segments in this dataset the mean segment duration is 1.66 s and the median is 1.05 s (interquartile range, IQR, 0.83–1.64 s). A single segment can therefore fall across two (or more) one-minute bins. Simply crediting the whole segment to the minute in which it starts would over-count that minute and under-count the next, so instead each segment's value is split across the minutes it covers, in proportion to how much of the segment falls in each minute:

$$q_b = q \cdot \frac{\min(t_{\text{end}}, b_{\text{end}}) - \max(t_{\text{start}}, b_{\text{start}})}{t_{\text{end}} - t_{\text{start}}}, \quad (5.4)$$

where q stands for any one of the four segment quantities listed above (child utterance count, child utterance duration, adult word count, and auditory-overlap duration), $[t_{\text{start}}, t_{\text{end}})$ is the segment's time span, and $[b_{\text{start}}, b_{\text{end}})$ is the minute it is being assigned to. Adding up these per-minute pieces across all segments gives, for each child and each minute, four acoustic features: total adult word count, total child utterance count, total child utterance duration, and total auditory-overlap duration. (The four features simply mirror the four segment quantities.) No measurement is lost or double-counted, since a segment's pieces always add back up to its original value. Segment lengths themselves vary, but because the split conserves the total this way, the per-minute features do not depend on where a segment's boundaries happen to fall relative to the minute boundaries.

5.4. Feature Fusion and Temporal Alignment

At this point there are two tables, both indexed by (child, minute): one of spatial features and one of acoustic features. Before joining them, all timestamps are converted to the recording site's local time zone so that the two streams line up. The tables are then merged on (child, minute), keeping only the minutes that appear in both, a *complete-case* join. Minutes for which one stream is missing (for example, a child has position data but the audio recorder was off that minute) are dropped, so that inclusion is never inferred from one modality when the other is unavailable.

5.5. Discovering Latent Activity Contexts

5.5.1. From Per-Child Behavior to a Room-Level View

Per-child sensor features are interpretable only in relation to the concurrent classroom activity. Ten minutes of peer-group time carries different meaning during whole-class instruction than during free play [43]. RQ1 therefore requires a transition from per-child measurement to a per-minute room-level view, allowing each one-minute epoch to be labelled with a shared activity context.

This room-level view is constructed by averaging each of four features across all children present in minute t ,

$$\bar{f}_t = \frac{1}{|C_t|} \sum_{i \in C_t} f_{i,t}, \quad (5.5)$$

where C_t is the set of children present in minute t and $f_{i,t}$ is child i 's value for feature f . The four features are adult word count, auditory overlap, cumulative displacement, and teacher distance. They are minimal sufficient under three constraints. (i) They cover the four axes the background literature identifies as the principal moderators of communicative demand in inclusive classrooms (§2.5): adult vocal input, acoustic demand on the listener, physical activity, and adult proximity. (ii) They are properties the children share at a given moment (room-level), making averaging across children meaningful. (iii) They exclude the per-child behavioural markers used as outcomes (time in a group, utterance count, group composition), avoiding circularity in the HL/TH comparison. Adding further candidates was considered and rejected on these grounds: number of children present is a function of attendance rather than activity; time of day would inject the routine's sequential structure into the cluster definition and create label leakage; pairwise inter-child distance is already absorbed by displacement and teacher distance. Minutes in which fewer than three children are jointly observed are

excluded from the room-level aggregation: with one child, the room average is that child's value; with two, it is the pair's mean. In both cases the "average" is indistinguishable from individual or pair-level behaviour, defeating the room-level construct. Three is the smallest count at which averaging across children produces a value distinct from any single child's behaviour.

5.5.2. Yeo–Johnson Transformation

Inspection of the four room-level feature distributions (Figure 5.1, left column) shows that three of the four are right-skewed: on the per-minute room averages, adult word count, cumulative displacement, and teacher distance have sample skewness of +1.13, +2.57, and +11.10 respectively, the last driven by occasional minutes in which the teacher is far from every child. Auditory overlap is close to symmetric (skewness -0.13). The features also differ in how heavy their tails are, that is, how often a minute takes a value far above the bulk of the distribution: this is by far most pronounced for teacher distance and then for cumulative displacement (excess kurtosis +178.8 and +13.5, against 0 for a normal distribution), mild for adult word count (+1.0), and essentially absent for auditory overlap (-0.4). Some features are also effectively zero-inflated; about 5% of minutes have near-zero adult speech (room-average AWC below one word per minute), consistent with quiet activity periods or moments when no adult is within range of the child's recorder.

A GMM assumes approximately Gaussian mixture components, so fitting directly on skewed, heavy-tailed features risks placing components along these distortions rather than along meaningful activity types.

A Yeo–Johnson power transformation [44] is applied per feature before fitting. Box–Cox is not an option here: two features can take values of zero (displacement when stationary, auditory overlap when silent), which Box–Cox does not handle. A fixed transform such as $\log(1 + x)$ would over- or under-correct the four features unequally, since their skews differ substantially (+1.13 to +11.10); a per-feature data-driven λ shrinks all four toward Gaussian shape simultaneously. Standardisation alone is insufficient because GMM component shape depends on the underlying distribution, not only its scale. After transformation, each feature is much closer to Gaussian (Figure 5.1, right column), with approximately zero skew and unit variance; the only remaining deviation is a small spike of near-zero adult-word-count minutes that no transformation of this kind can remove. Cluster profiles in Chapter 6 are re-expressed on the original feature scales for readability.

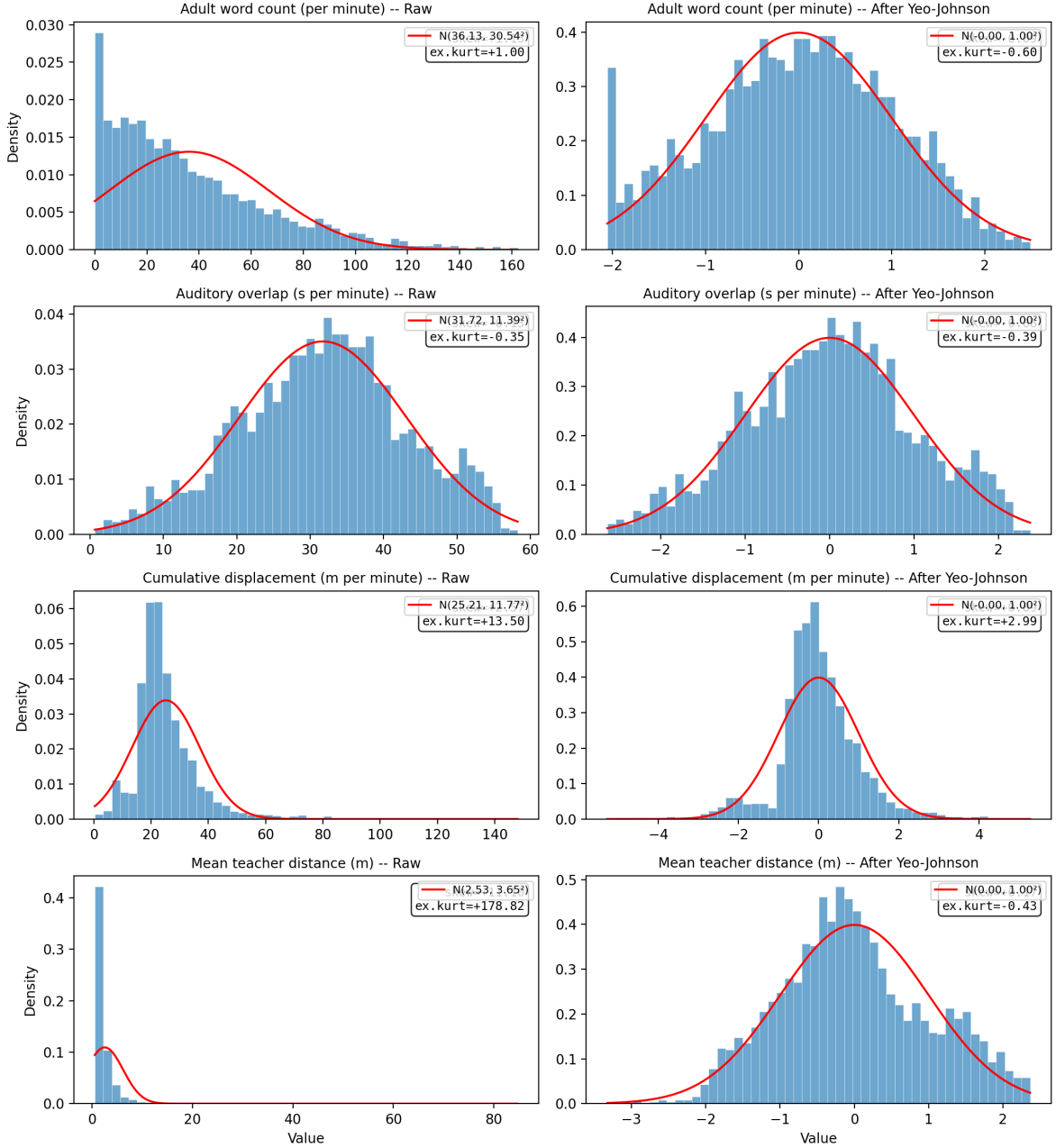


Figure 5.1: Marginal distributions of the four room-level clustering features, before (left) and after (right) the per-feature Yeo-Johnson transformation, over the per-minute room averages. A normal-distribution curve (red) is overlaid on each panel for reference, and each panel is annotated with its excess kurtosis (0 for a normal distribution; larger values indicate heavier tails). The raw adult word count, cumulative displacement, and teacher distance are right-skewed, with long upper tails (teacher distance especially so), while auditory overlap is roughly symmetric; after transformation all four are close to a standard normal distribution. The remaining spike at the left of the transformed adult-word-count panel comes from the many minutes with almost no adult speech, which a transformation of this kind cannot remove.

5.5.3. Gaussian Mixture Model

A GMM with full covariance represents the distribution of room-level features as

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (5.6)$$

where K is the number of mixture components (each later treated as one activity context), and component k has mixture weight π_k , mean μ_k , and full covariance matrix Σ_k . A full-covariance GMM is preferred to k -means because it produces soft posterior assignments (appropriate for gradual classroom transitions) and accommodates correlated feature directions (e.g. high AWC co-occurring with low displacement during whole-class instruction) that a spherical-covariance model would not [45]. The choice of covariance structure is empirically supported: at the selected $K = 6$ the full-covariance fit has mean BIC 21,203 against 21,792 for a diagonal-covariance fit and 22,692 for a spherical-covariance fit (a difference of $\Delta\text{BIC} \approx 589$ and $\approx 1,489$ respectively, decisively favouring the full-covariance model on the Kass–Raftery scale [46]). The full-covariance model uses 89 parameters at $K = 6$ against 53 and 35 for the diagonal and spherical variants; the BIC penalty for the extra parameters is more than offset by the gain in fit.

Although the GMM gives each minute a soft (probabilistic) membership in every component, each minute is then assigned to a single context by *hard assignment*: the component with the highest posterior probability for that minute. Throughout this thesis, “cluster” is shorthand for the set of minutes assigned to a given mixture component in this way, and “centroid” for that component’s mean μ_k . The per-cluster quantities reported in Chapter 6, including the percentage of the day spent in each cluster, are computed over these hard assignments. The cost of hard assignment is low here because cluster membership is generally confident: the median posterior probability of the assigned component is 0.90 (IQR [0.72, 0.98]), and only 5.9% of minutes have a maximum posterior below 0.5.

5.5.4. Choice of the Number of Components

The number of mixture components is selected by a *stability-aware BIC rule* over $K = 2$ –20. For each K in this range, twenty independent full-covariance GMMs are fit, each with ten random restarts and an independent random seed; the mean BIC and its standard deviation across seeds are recorded, together with the mean pairwise Adjusted Rand Index (ARI) between the twenty resulting label assignments. The ARI is a partition-agreement measure, ranging from 0 (random) to 1 (identical clustering up to label permutation); a high mean pairwise ARI at a given K indicates that different initialisations converge to the same partition, which is the empirical signature of a well-specified clustering rather than a fragile EM local optimum [47]. The selection rule has two steps: a value of K is *admissible* iff its mean pairwise ARI is at least 0.80, and the selected K is the admissible value with the lowest mean BIC. The 0.80 threshold is chosen so that label assignments must agree on at least roughly four-fifths of pair-wise classifications across seeds, well above chance and below saturation. The selected K is robust to this choice over a meaningful range: $K = 6$ is the lowest-BIC admissible value at any ARI threshold in [0.70, 0.85]; only at thresholds ≥ 0.86 does $K = 6$ drop out and $K = 5$ become selected. After fitting, cluster identifiers are re-indexed by ascending mean AWC, so that C_0 has the lowest AWC centroid and C_{K-1} the highest. This relabelling makes the IDs the same across runs and orders them along a single feature (AWC), from the lowest-AWC context to the highest.

Combining BIC with a stability requirement is necessary because BIC alone is known to be insufficient for choosing K in GMMs: when components are not strictly Gaussian, BIC frequently keeps finding shallow further minima at large K that reflect fitting noise rather than recovering structure [47, 45]. Standard practice therefore pairs BIC with an external stopping criterion (here, replication of the partition across random initialisations). This combined criterion replaces the more common “BIC argmin within a hand-chosen upper bound” that an earlier version of this analysis used; the upper bound is now set generously at $K = 20$ and is not where the selection rule does its work.

Figure 5.2 reports BIC and stability across the swept range. Mean BIC describes a broad basin

spanning $K = 5$ – 8 , with the across-seed standard deviation small and stable through this basin. From $K = 14$ onward the across-seed standard deviation grows sharply (over an order of magnitude larger than for $K \leq 13$), the empirical signature of fragile EM local optima at high K : at any single random restart a high- K fit may find a sharp local optimum that another seed will not reproduce, so the mean BIC at these K values masks a wide spread across runs. Mean pairwise ARI tracks this pattern: ARI is at or near 1 for $K = 2, 3, 5$, drops to 0.85 at $K = 6$ (still above the admissibility threshold), and falls below the 0.80 threshold at $K = 4$ and at every $K \geq 7$. The admissible set is therefore $\{2, 3, 5, 6\}$, and the lowest mean BIC within this set is at $K = 6$, which is selected. The slightly lower mean BIC at $K = 7$ (21,192 vs. 21,203 at $K = 6$) is not used, because the partition at $K = 7$ does not reproduce across seeds (mean pairwise ARI 0.69); the gap is also small on the Kass–Raftery scale of BIC differences [46], where the conventional thresholds for “positive”, “strong”, and “very strong” evidence are 2, 6, and 10 units respectively.

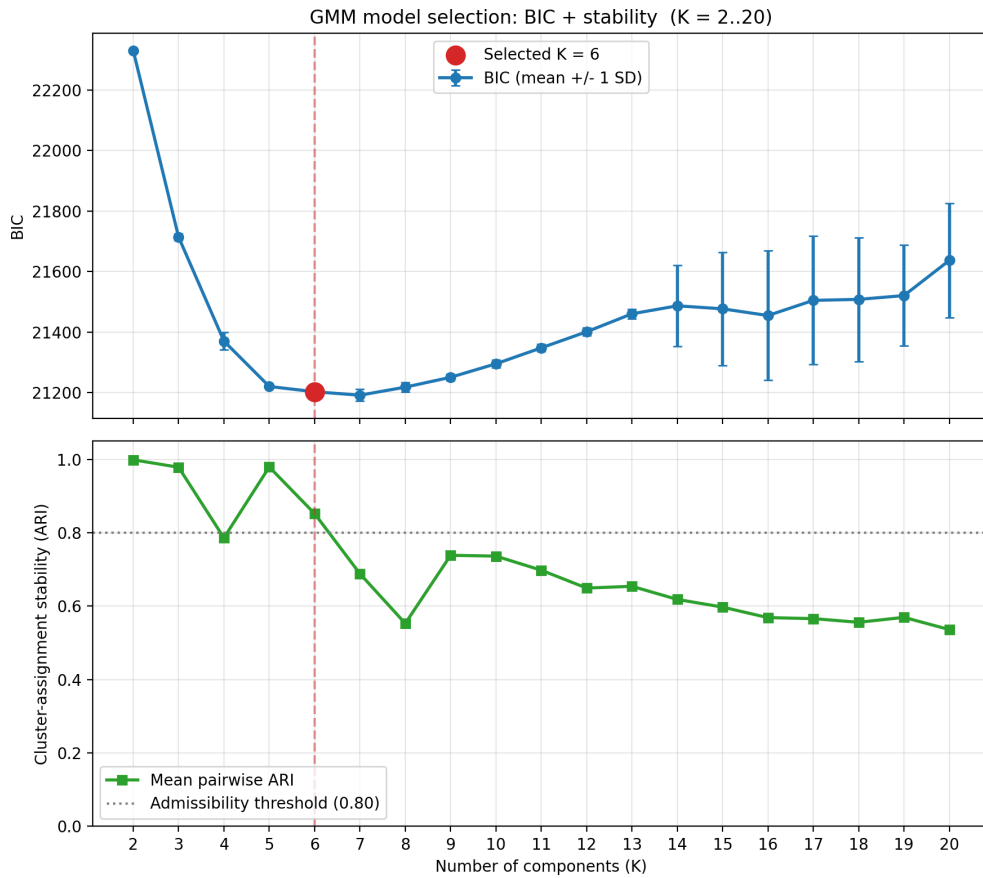


Figure 5.2: Stability-aware BIC sweep over $K = 2$ – 20 . Top: mean BIC across twenty independent random seeds at each K , with error bars at ± 1 standard deviation. Bottom: mean pairwise Adjusted Rand Index between the twenty resulting label assignments, with the admissibility threshold ($\text{ARI} \geq 0.80$) marked. The selected $K = 6$ (red marker) is the admissible value with the lowest mean BIC. From $K = 14$ onward the BIC standard deviation grows over an order of magnitude, the empirical signature of fragile EM local optima at high K .

The data-driven choice of $K = 6$ can be compared informally with the independent manual coding of the daily routine (Chapter 4), which identifies eight activity types. The two are close but do not coincide exactly. The BIC sweep is run blind to the manual coding, and the manual-coding count is consulted only after the data-driven K is fixed; cluster assignment and profile are recovered from the sensor data alone. Whether the six recovered components admit interpretable activity-type labels is

treated descriptively in §6.1.

5.6. Operationalising Social Inclusion per Child, per Context

With latent activity contexts available, each child’s behaviour is aggregated within each context to produce a context-conditional behavioural profile. The scope of the inclusion construct measured here warrants comment. The developmental literature defines inclusion in terms of belonging, acceptance, and response contingency [43, 3], none of which are directly measurable from UWB and LENA. What the sensors measure are *behavioural markers* of inclusion: spatial proximity, vocal output, and group composition. The phrase “behavioural markers of inclusion” is used throughout to signal that these are correlates of the underlying construct.

The three behavioural markers are as follows.

- **Peer co-presence.** The child’s percentage of each one-minute epoch spent in any spatial group. Operationalises the degree to which a child is spatially co-located in a peer group. It does not distinguish a child who is central to a group from one who is only on its edge.
- **Vocal participation rate.** The child’s utterance count per minute. Utterance count is preferred to duration because LENA’s count metric is the one most directly validated against human transcription for child vocalisations in naturalistic settings [33, 35]. This measure does not distinguish vocalisations that are responded to from those that are ignored.
- **Peer affiliation patterns.** The demographic composition of the child’s peer-group time, split into three percentages per minute, namely time in HL-only spatial groups, time in TH-only spatial groups, and time in mixed spatial groups. For within-diagnosis contrasts these are collapsed into a *same-diagnosis* outcome (time in HL-only spatial groups for an HL child, time in TH-only spatial groups for a TH child), which is the informative quantity since by definition an HL child is never in a TH-only group and vice versa. Of the three markers, peer affiliation is most directly tied to cross-group contact, a necessary but not sufficient precondition for sociometric inclusion.

The split between peer co-presence and vocal participation rate aligns with ethnographic observations of DHH children in inclusive classrooms [42]. HL children often achieve physical proximity to peers, for example by sitting close to hearing classmates or joining small-group activities, but participate less verbally in fast-paced multiparty exchange. The spatial group feature captures the former. The LENA vocalisation features capture the latter. Peer affiliation is a third axis, analysing the diagnostic composition of the group a child is in.

5.6.1. Context-Conditional Comparison Between HL and TH Children

After merging the per-child behavioural markers with the demographic mapping, HL and TH children are compared within each latent activity context. Conditioning on context is necessary because HL and TH children may be exposed to contexts in different proportions, which would bias an unconditional comparison. These context-conditioned comparisons form the input to the LMM described next.

5.7. Statistical Inference

5.7.1. Model

Each child contributes many minute-level observations, and observations from the same child are not independent of one another: a child who is generally sociable, or generally quiet, tends to look that way

across many of their minutes. An LMM accounts for this *within-child dependence* by giving each child their own random intercept [48]. For each of the four minute-level outcomes (peer co-presence, vocal participation rate, same-diagnosis group time, and mixed-diagnosis group time), the following model is fitted.

- **Fixed effects.** Cluster, hearing status (HL vs. TH), and their interaction.
- **Random effect.** Per-child intercept.

The random intercept absorbs stable between-child variation in the outcome, and the cluster $\times$ hearing-status interaction allows the HL/TH gap to differ across contexts. Random slopes for cluster within child were considered but not estimated: with 13 children and six clusters, each child contributes to only a subset of clusters in meaningful quantity, and a child $\times$ cluster slope structure would be under-identified. Random slopes for hearing status do not apply because hearing status is a between-subject variable. The chosen specification therefore identifies the cluster $\times$ hearing-status interaction with the minimum random structure required to absorb between-child dependence.

Same-diagnosis group time is the within-diagnosis contrast: time in HL-only spatial groups for an HL child, and time in TH-only spatial groups for a TH child (§5.6).

5.7.2. Per-Cluster Tests

Within each cluster, HL vs. TH is tested with a single-degree-of-freedom Wald contrast on the fitted LMM, combining the hearing-status coefficient with the relevant cluster-by-diagnosis interaction coefficient. Because each marker is tested across all six contexts, these contrasts are corrected for multiple comparisons (described next), and significance is judged on the corrected value: an effect is *significant* when its FDR-adjusted $q < 0.05$ and *marginal* when $0.05 \leq q < 0.10$. Raw p -values are reported alongside q throughout so both can be seen.

Effect size is reported as Cohen's d_s [49] computed from child-level means with a pooled within-group standard deviation,

$$d_s = \frac{\bar{M}_{HL} - \bar{M}_{TH}}{s_{\text{pooled}}}, \quad s_{\text{pooled}} = \sqrt{\frac{(n_{HL} - 1)s_{HL}^2 + (n_{TH} - 1)s_{TH}^2}{n_{HL} + n_{TH} - 2}}.$$

With $n_{HL} = 6$ and $n_{TH} = 7$ the small-sample bias in d_s is approximately 7% upward. Hedges' g correction was not applied because d is read as an indication of magnitude alongside the significance tests rather than as an exact estimate. Where $|d|$ is large (above 2 for some affiliation contrasts), it reflects HL and TH child-means that are clearly separated with little overlap, consistent with the significant tests for those contexts; for the homophily index and the exposure check, which are tested non-parametrically, d is reported as a descriptive companion to the rank-based result.

Across four outcomes and six contexts, these per-cluster contrasts amount to many tests (24 LMM contrasts, plus 6 exposure tests and 6 homophily tests, 36 in all). To guard against false positives from this multiplicity, a Benjamini–Hochberg false-discovery-rate (FDR) correction is applied within each marker family (the six within-context contrasts of a single outcome), and the resulting q -values are reported beside the raw p -values in every results table. The three markers operationalise three distinct constructs (spatial co-presence, vocal output, demographic composition of grouped time), each tested across the six contexts; correcting within each marker controls the expected proportion of false discoveries among the six within-context contrasts for that construct. Correcting across all 24 LMM contrasts jointly would conflate three independent hypotheses and inflate the type-II rate without

a corresponding benefit, particularly in an exploratory study with 13 children. A strict family-wise correction (such as Bonferroni) would be inappropriate for the same exploratory-aim reason: with only 13 children the aim is to find where HL/TH differences concentrate rather than to confirm each one in isolation. Benjamini–Hochberg is appropriate when the tests within a family are independent or positively dependent; the within-marker contrasts share the same fitted model and the same underlying social structure, which is a positive-dependence regime, so Benjamini–Yekutieli’s correction for arbitrary dependence would be unnecessarily conservative here. The 6 exposure tests are a balance check rather than hypotheses and are reported uncorrected (§5.7.4).

5.7.3. Homophily Index

The homophily index, the proportion of a child’s grouped time spent in same-diagnosis groups, is a per-child summary, one value per child per cluster, conditional on being in a spatial group. An index of 1 means a child is always in same-diagnosis groups when grouped, while 0 means always in mixed-diagnosis groups. Because it is a child-level rather than minute-level quantity, it is tested per cluster with a Mann–Whitney U test at child level ($N = 13$). Mann–Whitney U is used because the small sample (six HL, seven TH) cannot establish the normality assumption of a child-level t -test, and rank-based inference is robust to the moderately skewed child-level distributions observed for the exposure and homophily measures.

5.7.4. Exposure Check

Before interpreting within-cluster HL/TH differences, it is necessary to verify that HL and TH children are comparably exposed to each context. Per-cluster Mann–Whitney U tests on the child-level proportion of the day spent in each context ($N = 13$) check for *exposure asymmetry*, that is, HL and TH children spending systematically different amounts of time in a context. A large asymmetry would mean a within-context HL/TH comparison rested on very different amounts of data for the two groups.

6

Results

Results are reported in four parts. §6.1 presents the six latent activity contexts recovered by the GMM, with qualitative labels derived from their sensor profiles. §6.2 then checks exposure, confirming that HL and TH children are not grossly unevenly distributed across contexts. §6.3 shows the per-context behavioural markers visually, and §6.4 tests them formally with an LMM. Discussion is deferred to Chapter 7.

6.1. Discovered Latent Activity Contexts

The selected GMM has $K = 6$ components (§5.5.4). Each component groups together minutes with a similar room-level profile. To make them easy to refer to, the six contexts are numbered C_0 to C_5 in order of increasing mean adult word count (AWC), so C_0 has the lowest average adult speech and C_5 the highest. Each context's *sensor profile* is its average (centroid) value on the four room-level features (adult word count, auditory overlap, cumulative displacement, and teacher distance), shown in Figure 6.1.

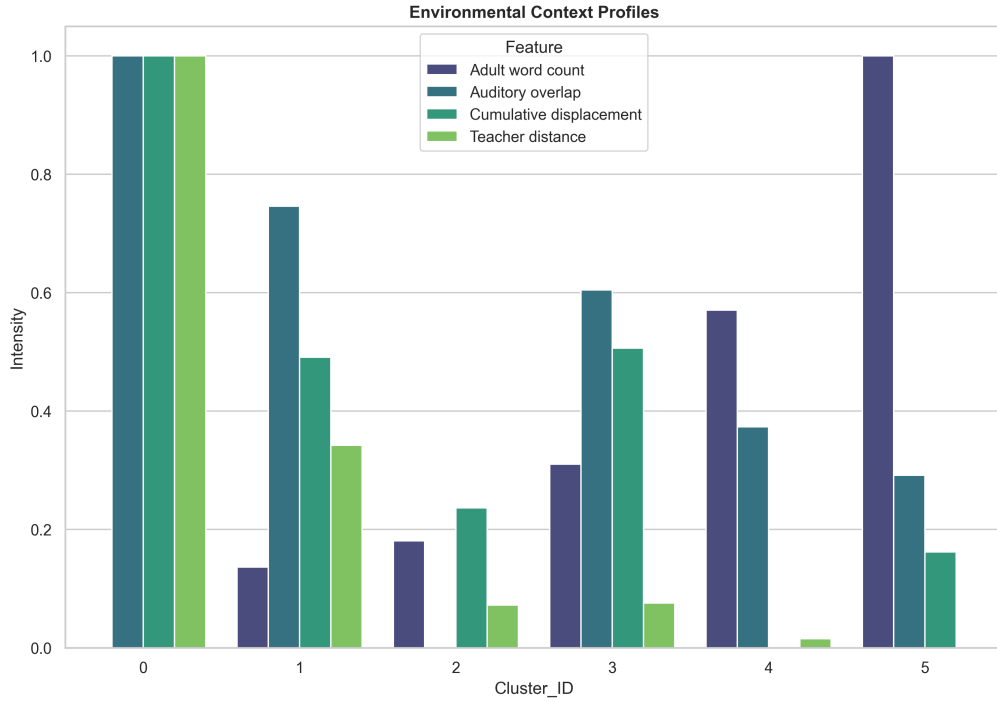


Figure 6.1: Sensor profile of each of the six contexts: the average (centroid) value of the four room-level features. Because the four features are measured in different units, each feature is rescaled to $[0, 1]$ across the six contexts (min-max scaling) so they can be compared on one axis; a bar of height 1.0 marks the context with the highest average for that feature and 0.0 the lowest. The features used here are only the four room-level clustering features, not the behavioural outcomes.

Figure 6.1 rescales each feature to $[0, 1]$ for visual comparison. Table 6.1 gives the same centroids in their original units, so the absolute level of each feature can be read directly.

Table 6.1: Per-cluster centroids of the four room-level features in their original (unscaled) units: adult word count per minute, auditory-overlap duration in seconds per minute, cumulative displacement in metres per minute, and mean teacher distance in metres. Each value is the mean of the per-minute room average (Eq. 5.5) over the minutes assigned to that cluster; Figure 6.1 shows the same profiles after min-max rescaling.

Cluster	AWC (words/min)	Overlap (s/min)	Displacement (m/min)	Teacher dist. (m)
0	2.22	51.57	41.73	9.35
1	10.95	42.17	28.75	4.18
2	13.77	14.55	22.25	2.06
3	22.02	36.93	29.14	2.09
4	38.65	28.37	16.22	1.61
5	66.06	25.34	20.35	1.49

Each context is described below from its sensor profile in Figure 6.1, together with the per-cluster behavioural statistics reported later in this chapter. The words “low”, “moderate”, and “high” are used loosely, to say where a context sits relative to the others on a feature; the exact ordering by AWC is the cluster numbering itself.

Two kinds of quantity appear in these descriptions, and it is important to keep them separate. The four room-level features (adult word count, auditory overlap, displacement, teacher distance) are what the GMM used to form the contexts, and these are the values plotted in Figure 6.1. The behavioural markers (utterance counts and time spent in spatial groups) were deliberately excluded from clustering (§5.5); they are mentioned in the descriptions only to help characterise each context after it has been

formed, and play no part in forming it.

Cluster 0: Dispersed transition.

AWC is the lowest of any cluster, paired with high auditory overlap, the highest cumulative displacement (41.7 m/min), and the highest teacher distance (~9.4 m). Peer co-presence is the lowest of any cluster (~20–23%), so children are moving across the room rather than gathered into spatial groups. The high auditory overlap here need not indicate peer conversation, since the auditory-overlap feature also captures background noise (§5.3). The combination of high movement, wide dispersion, low adult speech, and low peer co-presence is the sensor signature of a transition rather than of an active peer-interaction phase; the label is descriptive in that sense, not a claim that any one routine event (e.g. between-activity movement, arrival, clean-up) is being recovered. It accounts for ~6.9% of observed time.

Cluster 1: Peer-driven activity.

This context combines low AWC and high auditory overlap with moderate-to-high displacement and a moderately high teacher distance (~4.2 m); children are moving and vocalising while the teacher is comparatively far away. Utterance counts are the highest of any cluster (~4.0 per minute for TH children). It accounts for ~6.0% of observed time. The same sensor profile could equally arise from other configurations such as transitions or outdoor periods.

Cluster 2: Independent / parallel work.

AWC is low and auditory overlap the lowest of any cluster (the room is quiet), with displacement and teacher distance both low-to-moderate. Utterance counts are the lowest of any cluster (~1.3–1.7 per minute), consistent with individual or small-group quiet work alongside a present-but-quiet adult. It accounts for ~8.6% of observed time.

Cluster 3: Adult-scaffolded peer activity.

Moderate AWC sits alongside high auditory overlap, moderate-to-high displacement, and low teacher distance. Peer speech dominates the acoustic environment (high overlap) while a teacher remains close (~2.1 m) but produces less speech than during direct instruction. This is the second-largest cluster by share of observed time (~33.9%), with both groups in spatial groups ~40% of the time and utterance counts among the highest in the dataset (~2.7–3.6 per minute), consistent with child-driven cooperative activity (centres, small-group play, snack time) with adult supervision nearby. The label refers to teacher proximity combined with peer-dominated speech, not to scaffolding in the interactional sense of contingent uptake [28], which is not measured by the sensors used here.

Cluster 4: Seated guided work.

AWC is high and displacement is the lowest of any cluster (~16 m per minute), with teacher distance the second-lowest (~1.6 m) and auditory overlap moderate. Peer co-presence is high (~40–46%), consistent with a seated, teacher-adjacent format (table activities, structured small-group work) in which children are physically still and the teacher is closely guiding the task. It accounts for ~9.8% of observed time.

Cluster 5: Whole-class instruction / read-aloud.

AWC is the highest of any cluster (by a wide margin: ~66 words per minute), auditory overlap is moderate, displacement is low, and teacher distance is the lowest of any cluster (~1.5 m). This is the most adult-vocally-dominated moment of the day. Utterance counts are below average

(~2.2–2.5), consistent with children listening rather than producing. This is the largest cluster by share of observed time (~34.8%).

These labels are based on the sensor profiles in Figure 6.1, not carried over from any pre-existing scheme. Each of the six contexts has a distinct profile (no two are near-duplicates, and none is empty or extreme), and each can be matched to a recognisable phase of an inclusive preschool day; these phases are qualitatively consistent with the manually coded activity types of Chapter 4, although no quantitative comparison (e.g. contingency analysis between cluster assignments and manual codes) was performed. The recovered structure separates two pairs of contexts that differ along a single dimension: C_0 and C_1 are both low-AWC, low-adult-presence settings that differ in dispersion (C_0 with children spread across the room, C_1 with children clustered into a peer-driven group); C_3 and C_5 are both teacher-proximal settings that differ in whether peer or adult speech dominates (C_3 with children speaking and the teacher monitoring, C_5 with the teacher speaking and the children listening). The remaining two contexts cover quiet individual work (C_2) and seated teacher-guided activity (C_4). These descriptive correspondences are suggested by the sensor profiles, not validated against an independently coded ground truth. The statistical conclusions in §6.4 are computed on cluster identifiers (C_0 to C_5), so reading every reported contrast with cluster numbers in place of labels leaves the test statistics unchanged. The labels are interpretive aids for the discussion, not inputs to the inference.

6.2. Differential Exposure Across Contexts

Before comparing HL and TH children within each context, it is worth checking that the two groups spend roughly comparable amounts of time in each one. This matters because if one group almost never entered a given context, the HL/TH comparison there would rest on very few minutes and would be unreliable. Using the hard cluster assignments defined in §5.5, Table 6.2 reports the mean percentage of each child’s observed day spent in each context, by hearing status.

Table 6.2: Mean percentage of observed day spent in each cluster, by hearing status. The p column reports the per-cluster Mann–Whitney U test on child-level proportions.

Cluster	HL mean (%)	TH mean (%)	p
0	7.3	6.6	0.445
1	6.1	6.1	1.000
2	8.2	9.0	0.366
3	34.5	33.1	0.295
4	9.1	10.5	0.836
5	34.8	34.6	1.000

No per-cluster Mann–Whitney test reaches the conventional significance threshold, and no cluster shows even a marginal difference. The largest absolute gap is in Cluster 4 (~1.5% of the day, $p = 0.836$). With $N = 13$ the test has limited power to resolve small differences, so this should not be read as positive evidence of equal exposure but as the absence of a large asymmetry. HL and TH children are comparably exposed to the six contexts overall, and within-context comparisons rest on meaningful per-cell child counts.

Table 6.3 reports the absolute per-cluster minute counts split by hearing status, so the reader can judge the fragility of any individual contrast directly. The smallest per-cell HL count is in Cluster 1 ($N_{HL} = 602$ minutes), and the smallest TH count is also in Cluster 1 ($N_{TH} = 717$); the largest cells are Clusters 3 and 5, each contributing roughly a third of the dataset.

Table 6.3: Per-cluster minute counts contributing to the LMM contrasts, split by hearing status. “Total” is the column total used in the percentages of Table 6.2. Across the 21,847 attendance-confirmed minutes, 9,938 come from HL children and 11,909 from TH children.

Cluster	N_{HL} minutes	N_{TH} minutes	Total
0	726	787	1,513
1	602	717	1,319
2	815	1,065	1,880
3	3,421	3,974	7,395
4	901	1,231	2,132
5	3,473	4,135	7,608
Total	9,938	11,909	21,847

6.3. Behavioural Markers by Context and Hearing Status

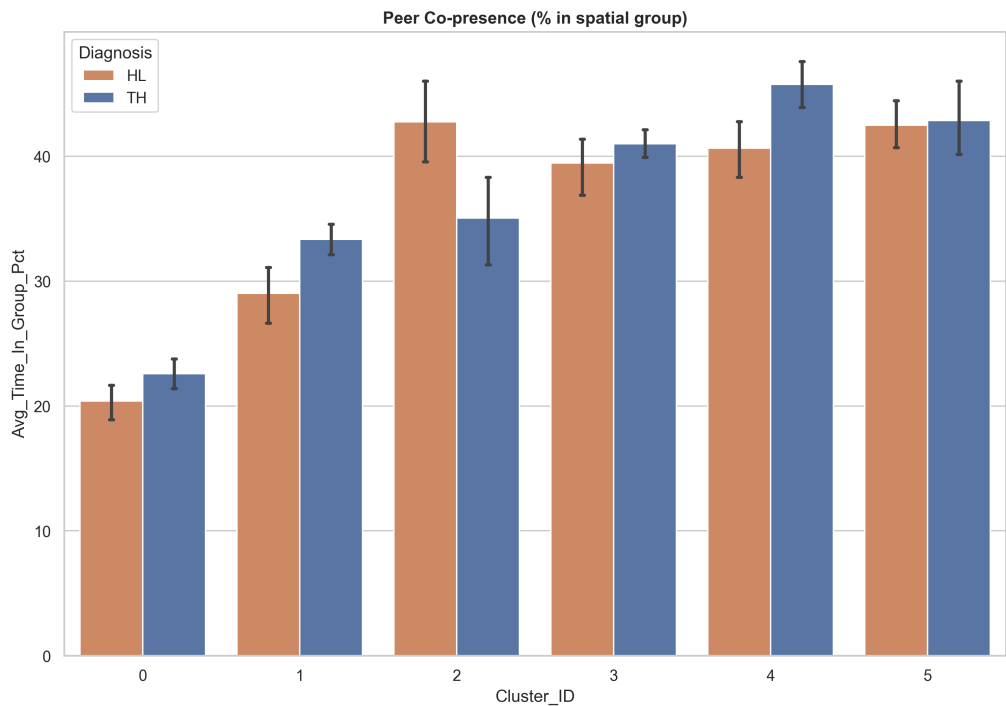


Figure 6.2: Peer co-presence (percentage of each one-minute epoch spent in any spatial group) by cluster and hearing status. Error bars are 95% confidence intervals over child means.

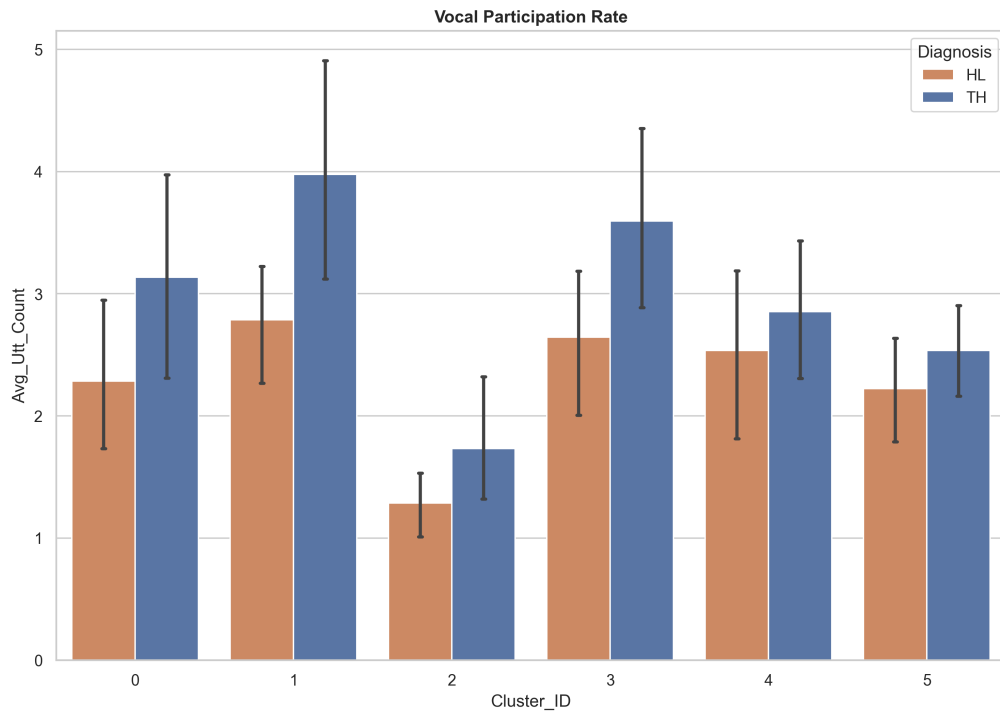


Figure 6.3: Vocal participation rate (mean child utterance count per minute) by cluster and hearing status. Error bars are 95% confidence intervals over child means.

Figure 6.2 shows three visible HL/TH differences in peer co-presence. In Cluster 2 (independent / parallel work), HL children spend more time in spatial groups than TH children. In Cluster 1 (peer-driven activity) and Cluster 4 (seated guided work), HL children spend less time in spatial groups than TH children. In the other three contexts the two groups track each other closely. Figure 6.3 shows a consistent tendency for TH children to vocalise more than HL children, with the visible gap largest in Clusters 1 and 3 (peer-driven activity and adult-scaffolded peer activity) and visible but smaller in Cluster 0 (dispersed transition).

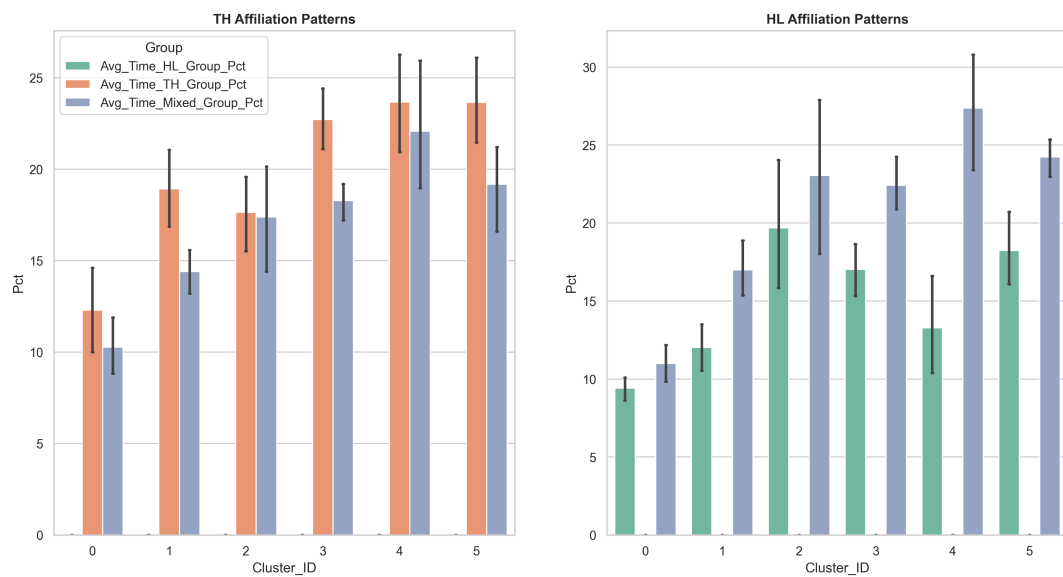


Figure 6.4: Peer affiliation patterns by cluster. Left, mean percentage of TH children's grouped time spent in HL-only, TH-only, and mixed spatial groups. Right, the same for HL children. Error bars are 95% bootstrap confidence intervals over children.

Figure 6.4 shows a clear asymmetry in peer affiliation. TH children spend most of their grouped time in TH-only spatial groups, with a smaller contribution from mixed-group time and essentially none in HL-only groupings. HL children show a more balanced split. In several contexts (Clusters 1, 3, and 4) HL children spend more grouped time in mixed-diagnosis spatial groups than HL-only ones. This asymmetry, in which TH grouped time is concentrated in same-diagnosis groupings while HL grouped time is more evenly distributed, holds across the six contexts and is developed in the LMM analyses that follow.

6.4. Linear Mixed Model Findings

The LMM uses child-minute observations with a per-child random intercept. Pooled across clusters at child-mean level, HL/TH differences are small for peer co-presence (HL 38.8%, SD 2.2; TH 39.9%, SD 2.5) and moderate for vocal participation rate (HL 2.4, SD 0.6; TH 3.0, SD 0.7 utterances per minute), where each standard deviation is taken across children. The hearing-status signal, where present, is concentrated in specific cluster-by-diagnosis interactions rather than in a uniform across-context contrast.

6.4.1. Peer Co-presence

The LMM on minute-level peer co-presence finds per-cluster Wald contrasts that survive FDR correction in three contexts: Cluster 1 (peer-driven activity, HL below TH), Cluster 2 (independent / parallel work, HL above TH), and Cluster 4 (seated guided work, HL below TH). Within-cluster Wald tests are in Table 6.4.

H1 predicted no consistent HL deficit across contexts, with HL children sustaining or exceeding TH co-presence where communicative demand is relaxed. The observed pattern is consistent with this in three of six contexts (no significant HL/TH gap in Clusters 0, 3, and 5) and in the direction H1 predicts in Cluster 2, where HL children exceed TH in independent / parallel work, a low-demand context. The remaining two deviations qualify H1. The Cluster 1 deviation (HL below TH in peer-driven activity) sits in a high-demand context that H1's compensatory account did not directly address. The Cluster 4 deviation (HL below TH in seated guided work) is the clearer contradiction: this is an adult-mediated context in which H1 would predict HL co-presence to meet or exceed TH, and the data instead show HL below.

Table 6.4: Child-level mean peer co-presence (% of minute in any spatial group) per cluster, by hearing status. “HL–TH” and “within-cluster p ” are the LMM Wald contrast estimate and test respectively. Because these are estimated jointly with the random intercept, they differ slightly from the raw HL–TH difference of the child-mean columns, and where the effect is close to zero the two can even differ in sign. Cohen's d [49] is computed at child-mean level. The q (BH) column is the Benjamini–Hochberg false-discovery-rate-adjusted p -value within this marker's six contexts. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Cluster	HL (%)	TH (%)	HL–TH	d	p	q (BH)	Context
0	20.40	22.57	–1.74	–1.19	0.284	0.341	Dispersed transition
1	29.01	33.33	–4.32	–1.84	0.010*	0.020*	Peer-driven activity
2	42.75	35.04	+7.84	+1.59	< 0.0001***	< 0.0001	Independent / parallel work
3	39.45	41.01	–1.47	–0.63	0.277	0.341	Adult-scaffolded peer activity
4	40.64	45.74	–5.33	–1.65	0.0005***	0.002**	Seated guided work
5	42.48	42.86	–0.72	–0.11	0.595	0.595	Whole-class instruction

Three within-cluster comparisons survive FDR correction. In Cluster 2 (independent / parallel

work), HL children exceed TH by about 7.8% of the minute ($p < 0.0001$, $q < 0.0001$, $d = +1.59$). In Cluster 4 (seated guided work), HL children fall below TH by about 5.3% of the minute ($p = 0.0005$, $q = 0.002$, $d = -1.65$). In Cluster 1 (peer-driven activity), HL children fall below TH by about 4.3% of the minute ($p = 0.010$, $q = 0.020$, $d = -1.84$). The three effects each rest on ~6–10% of the observed day, so their magnitudes should be read in light of the small per-cluster sample base.

6.4.2. Vocal Participation Rate

The LMM on minute-level utterance count finds per-cluster Wald contrasts that reach significance at Clusters 1 and 3 (peer-driven activity and adult-scaffolded peer activity); Cluster 0 (dispersed transition) runs in the same direction but is only marginal after correction. Within-cluster Wald tests are in Table 6.5.

Table 6.5: Child-level mean vocal participation rate (child utterance count per minute) per cluster, by hearing status. Columns as in Table 6.4. * $p < 0.05$, ** $p < 0.01$.

Cluster	HL	TH	HL–TH	d	p	q (BH)	Context
0	2.28	3.13	–0.86	–0.81	0.035*	0.071	Dispersed transition
1	2.79	3.97	–1.14	–1.05	0.006**	0.034*	Peer-driven activity
2	1.29	1.73	–0.46	–0.73	0.254	0.381	Independent / parallel work
3	2.65	3.60	–0.92	–0.96	0.016*	0.048*	Adult-scaffolded peer activity
4	2.53	2.85	–0.23	–0.37	0.562	0.562	Seated guided work
5	2.22	2.54	–0.35	–0.54	0.368	0.442	Whole-class instruction

HL children produce fewer utterances per minute than TH peers in Clusters 1 and 3 (peer-driven activity $p = 0.006$, $q = 0.034$, $d = -1.05$; adult-scaffolded peer activity $p = 0.016$, $q = 0.048$, $d = -0.96$; both surviving Benjamini–Hochberg correction), and in the same direction in Cluster 0 (dispersed transition, $p = 0.035$, $d = -0.81$), though this effect is only marginal after correction ($q = 0.071$). In the three remaining contexts the difference is in the same direction but smaller and not significant. Examined post-hoc, the three contexts in which the gap is significant or marginal share a common feature profile: low-to-moderate AWC combined with relatively high peer activity (high overlap, high movement, or both). This shared profile emerges from inspection rather than from a pre-registered structural test, and the small number of clusters (6) limits the interpretive weight of the pattern. Cluster 0 draws on only ~6.9% of the day and has the lowest peer co-presence of any cluster, so its effect is best read as consistent with Clusters 1 and 3 rather than as an independent confirmation of equal weight.

6.4.3. Peer Affiliation Patterns

Affiliation asks not just whether HL and TH children are co-present and vocalising but with whom. Since an HL child is by definition never in a TH-only spatial group (and vice versa), the comparisons are HL-in-HL-only time against TH-in-TH-only time (same-diagnosis contrast) and HL-in-mixed against TH-in-mixed (mixed-diagnosis contrast), within each cluster.

The baseline rates produced by cohort composition alone require comment before the tables. With 6 HL and 7 TH children, a TH child has 6 same-diagnosis peers out of 12 classmates (a proportion of 0.50), while an HL child has 5 out of 12 (about 0.42). Under purely random dyadic affiliation the TH same-diagnosis index would therefore sit about 0.08 above the HL one, even with no preference at all. The HL/TH contrasts below should be read with this baseline asymmetry in mind. It is revisited quantitatively at the end of this section and in §7.3.

Same-diagnosis group time. Across five of six contexts, TH children spend more time in TH-only spatial groups than HL children spend in HL-only spatial groups (Table 6.6), significantly in four (Clusters 1, 3, 4, and 5). The direction reverses only in Cluster 2 (independent / parallel work), where the reversal is small and marginal.

Table 6.6: Child-level mean same-diagnosis group time per cluster. For HL children this is time in HL-only spatial groups. For TH children it is time in TH-only spatial groups. “HL–TH” and “within-cluster p ” are from the LMM Wald contrast. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Reported d values are uncorrected d_s (§5.7); the small-sample Hedges’ correction would shrink the magnitudes by approximately 7%.

Cluster	HL (%)	TH (%)	HL–TH	d	p	q (BH)	Context
0	9.41	12.30	–2.38	–1.14	0.109	0.109	Dispersed transition
1	12.02	18.92	–7.14	–2.74	< 0.0001***	< 0.0001	Peer-driven activity
2	19.70	17.64	+2.70	+0.47	0.059	0.071	Independent / parallel work
3	17.03	22.72	–5.61	–2.39	< 0.0001***	< 0.0001	Adult-scaffolded peer activity
4	13.28	23.67	–11.72	–2.50	< 0.0001***	< 0.0001	Seated guided work
5	18.24	23.67	–5.60	–1.64	< 0.0001***	< 0.0001	Whole-class instruction

Mixed-diagnosis group time. The complementary pattern appears in Table 6.7. Across all six clusters, HL children spend more time in mixed-diagnosis spatial groups than TH children do, significantly in four (Clusters 2–5); the Cluster 1 difference (+2.8% of the minute) is marginal ($p = 0.051$, $q = 0.061$). HL–TH differences range from +2.8 to +6.4% of the minute across the five clusters other than the dispersed-transition context, the largest in Cluster 4 (seated guided work, +6.4% of the minute), the same context in which HL children’s overall peer co-presence and same-diagnosis group time are most depressed.

Table 6.7: Child-level mean mixed-diagnosis group time per cluster. Positive HL–TH values indicate HL children spending more of their minute in mixed spatial groups than TH children do. “HL–TH” and “within-cluster p ” are from the LMM Wald contrast. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Cluster	HL (%)	TH (%)	HL–TH	d	p	q (BH)	Context
0	10.98	10.27	+0.64	+0.36	0.648	0.648	Dispersed transition
1	16.99	14.41	+2.83	+1.20	0.051	0.061	Peer-driven activity
2	23.05	17.40	+5.14	+1.07	0.0001***	0.0003	Independent / parallel work
3	22.42	18.29	+4.15	+2.20	0.0002***	0.0004	Adult-scaffolded peer activity
4	27.36	22.07	+6.39	+1.02	< 0.0001***	< 0.0001	Seated guided work
5	24.24	19.19	+4.89	+1.70	< 0.0001***	< 0.0001	Whole-class instruction

Homophily index. To combine the same- and mixed-diagnosis pictures into a single measure, the homophily index (defined in §5.6) is computed on each child’s grouped-minute totals within each cluster, conditional on being in a spatial group. Per-cluster Mann–Whitney U tests are in Table 6.8.

Table 6.8: Per-child, per-cluster homophily index, the proportion of grouped time spent in same-diagnosis groups, conditional on being in a spatial group. Mann–Whitney U compares HL and TH children at child level. $*p < 0.05$, $**p < 0.01$. Reported d values are uncorrected d_s (§5.7); the small-sample Hedges’ correction would shrink the magnitudes by approximately 7%.

Cluster	HL idx	TH idx	HL–TH	d	p	q (BH)	Context
0	0.46	0.54	−0.08	−0.82	0.234	0.281	Dispersed transition
1	0.41	0.57	−0.15	−2.48	0.002**	0.007**	Peer-driven activity
2	0.46	0.51	−0.04	−0.42	0.534	0.534	Independent / parallel work
3	0.43	0.55	−0.12	−2.90	0.001**	0.007**	Adult-scaffolded peer activity
4	0.33	0.52	−0.19	−1.90	0.022*	0.033*	Seated guided work
5	0.43	0.55	−0.13	−1.92	0.005**	0.009**	Whole-class instruction

The homophily index is lower for HL than TH in all six contexts, significantly in Clusters 1, 3, 4, and 5 (peer-driven activity, adult-scaffolded peer activity, seated guided work, and whole-class instruction). The direction never reverses. A natural reading is that TH children preferentially affiliate with same-diagnosis peers while HL children’s grouped time is more evenly split. Quantitatively, however, the observed gap across the four significant clusters (0.12–0.19 on the 0–1 scale) exceeds the ~0.08 baseline gap predicted by random dyadic affiliation by 0.04–0.11. Preference accounts for part of the asymmetry and baseline composition for a substantial share. The pattern is consistent with findings that typically developing children in mainstream inclusive classrooms often show same-group affiliation preferences [43, 27].

Discussion

7.1. Findings by Research Question

RQ1. Each of the six components has a distinct sensor profile and maps to a recognisable phase of the inclusive preschool day. Two pairs differ along a single dimension: C_0 and C_1 along physical dispersion, and C_3 and C_5 along whether adult or peer speech dominates (§6.1). The answer to RQ1 is therefore affirmative, subject to the labelling caveats noted in §6.1. Whether the six-context decomposition is specific to this classroom is taken up in §7.3 and Chapter 10 below.

RQ2. The LMM does not yield a uniform HL/TH effect. Per-cluster Wald contrasts reach significance in different, context-specific subsets and vary substantially in magnitude, consistent with context-dependent rather than uniform effects.

For peer co-presence, three contexts carry the interaction with opposing signs. In Cluster 2 (independent / parallel work), HL children are in spatial groups more than TH peers ($HL-TH = +7.8\%$ of the minute, $d = +1.59$, $q < 0.0001$). The natural reading of this effect is not obvious. The compensatory-presence pattern anticipated in §2.4 is one possibility, but it fits awkwardly here: the compensatory account explains HL children sustaining co-presence when a fast verbal peer channel is closed to them, whereas Cluster 2 is a quiet work context in which there is little such channel to be excluded from. Plausible alternative accounts are at least as strong: teacher-directed seating that places HL children close to peers during quiet work, activity-area composition that allocates HL children disproportionately to the table-based activities that fall into this cluster, or a propensity of HL children to seat themselves near peers for visual access to demonstrated tasks. The available data do not distinguish among these accounts; seating-position records or activity-area labels would be required to do so. In Cluster 1 (peer-driven activity) and Cluster 4 (seated guided work), the gap reverses sign. In peer-driven activity, HL children fall below TH by approximately 4.3% of the minute ($d = -1.84$, $q = 0.020$), consistent with the same modality-mismatch reasoning at the opposite end of the demand range, on the assumption that rapid peer talk increases the difficulty of group entry. The largest HL-below-TH co-presence gap is in Cluster 4 (seated guided work, $HL-TH = -5.3\%$ of the minute, $d = -1.65$, $q = 0.002$). The Cluster 2 effect is larger in absolute magnitude (+7.8% of the minute) but is discussed separately above because its sign reverses and its interpretation does not rest on the modality-mismatch framing. The same context (Cluster 4) shows the largest mixed-diagnosis gap (Table 6.7) and the largest same-diagnosis gap

(Table 6.6): TH children's grouped time during seated guided work is concentrated in TH-only groups, and HL children participate in these groups less frequently.

The Cluster 2, Cluster 1, and Cluster 4 effects each draw on ~6–10% of the observed day, so their magnitudes should be interpreted with caution and may partly reflect the specific organisation of this classroom during those minutes rather than a generalisable pattern.

H1 comprises two claims, which the data support unevenly. The first claim, that peer co-presence will not show a consistent HL deficit across contexts, depends on which criterion is used. By raw direction the gap is HL-below-TH in five of six contexts (Clusters 0, 1, 3, 4, and 5) and HL-above-TH only in Cluster 2, so the deficit is directionally consistent. By the stricter criterion of statistically significant per-cluster differences after FDR correction, the gap takes opposite signs (HL above TH in Cluster 2, HL below TH in Clusters 1 and 4) and is not significant in the remaining three contexts, which is the context-specific rather than uniform pattern H1 anticipated. The magnitudes vary substantially across contexts under either reading. H1's first claim is therefore supported in its significance-based reading but not in its directional reading; the small non-significant HL-below-TH gaps in Clusters 0, 3, and 5 should be acknowledged rather than read as null. The second claim, that where differences exist HL children sustain or exceed TH co-presence in contexts of relaxed communicative demand, is supported in Cluster 2 (low auditory overlap and low adult speech, with HL exceeding TH) and contradicted in Cluster 4 (an adult-mediated seated context in which the compensatory account would predict HL co-presence to meet or exceed TH, whereas HL falls below TH). H1 said nothing about high-demand contexts, so the Cluster 1 deviation (HL below TH in peer-driven activity) is not adjudicated by H1; it is consistent with extending the modality-mismatch framing to the high-demand end of the range, but that extension is post-hoc and is read here as a question for replication rather than as a confirmation of H1.

For vocal participation rate, the pattern is more coherent. HL children produce fewer utterances per minute than TH peers in the two contexts that combine high peer communicative demand with reduced adult scaffolding, peer-driven activity ($p = 0.006$, $q = 0.034$, $d = -1.05$) and adult-scaffolded peer activity ($p = 0.016$, $q = 0.048$, $d = -0.96$), and in the same direction in dispersed transition ($p = 0.035$, $d = -0.81$), which is only marginal after correction ($q = 0.071$). Reduced scaffolding manifests differently in each. In Cluster 3 (adult-scaffolded peer activity) it arises from moderate AWC combined with high auditory overlap, in which peer speech dominates the acoustic environment despite a teacher remaining close. In Cluster 1 (peer-driven activity) it arises from low AWC combined with high teacher distance. In Cluster 0 (dispersed transition) it arises from minimum AWC combined with high movement and overlap and the largest teacher distance of any cluster. In the three contexts where the gap is small and not significant, the demand on rapid peer turn-taking is attenuated either by very low auditory overlap (Cluster 2, independent / parallel work) or by adult speech dominating the acoustic environment (Clusters 4 and 5, seated guided work and whole-class instruction).

H2 is supported in direction but its operational axis required post-hoc revision. The original formulation framed communicative demand mainly in terms of adult speech volume and adult mediation of turn-taking; the data from this cohort instead point to the balance of peer-versus-adult vocal activity, with teacher distance not appearing to be the operative moderator. Cluster 3 makes this revision necessary: it has only moderate AWC and a close teacher, conditions under which the original H2 would predict a small gap, but its auditory overlap exceeds its AWC (a peer-dominated acoustic environment), and a significant gap is observed there alongside Cluster 1 where AWC is lower and the teacher is more distant. The gap is uniformly in the HL-below-TH direction across all six contexts, so the interaction reflects variation in magnitude rather than in sign. The revised reading of H2 (operative axis = peer-vs-adult balance) is a hypothesis suggested by the present data rather than a confirmation;

whether it generalises beyond this cohort is a question for replication.

For peer affiliation patterns, the HL/TH asymmetry is the most consistent of the three. TH children concentrate their grouped time in TH-only spatial groups more than HL children concentrate theirs in HL-only spatial groups in five of six contexts (significantly in four). HL children spend more time in mixed spatial groups than TH children do in all six contexts (significantly in four). The homophily index is lower for HL than TH in all six clusters, significantly in four. The direction never reverses (same-diagnosis time reverses marginally in Cluster 2, but the homophily index does not). Cohort composition (6 HL vs. 7 TH) accounts for roughly 0.08 of the homophily-index gap, so of the observed 0.12–0.19 in the four significant clusters, 0.04–0.11 reflects preference beyond what availability alone would predict.

H3 also made a second prediction: that this asymmetry would be larger in contexts in which children choose their own groupings than in teacher-organised ones, on the reasoning that homophily is stronger when group composition is self-selected [15]. Among the recovered clusters, only Cluster 1 (peer-driven activity) sits clearly in the child-chosen category; Clusters 4 (seated guided work) and 5 (whole-class instruction) sit clearly in the teacher-organised category; Cluster 0 (dispersed transition) is typically adult-cued and is not classed here as child-chosen. The data do not support the second prediction. The homophily-index gap is significant in the one child-chosen context (Cluster 1) but is equally significant in both teacher-organised contexts (Clusters 4 and 5), and the largest gap (Cluster 4) is itself in a teacher-organised context. Same-status grouping is therefore not confined to child-selected settings and persists when the activity is adult-organised. This pattern may reflect affiliation preferences from free-choice contexts carrying over into teacher-led configurations, although with 13 children and no experimental manipulation of group composition, this study cannot distinguish that explanation from alternatives.

H3 is therefore supported in its first part with two qualifications: the predicted direction holds in every context on the homophily index and the mixed-diagnosis measure, and in five of six contexts on same-diagnosis time (Cluster 2 reverses marginally, $p = 0.059$). The second part of H3, the predicted moderation by context type, is not supported.

The three markers differ in how their HL/TH gaps distribute across contexts. Peer co-presence varies by context: HL children are below TH in peer-driven activity (Cluster 1) and seated guided work (Cluster 4), above TH in independent / parallel work (Cluster 2), and statistically indistinguishable from TH in the other three contexts. HL children vocalise less in contexts where rapid multiparty peer talk dominates the acoustic environment, whether the teacher is distant (Cluster 1) or close but producing less speech (Cluster 3). Independent of context, TH grouped time is disproportionately same-diagnosis while HL grouped time is more evenly distributed — the only one of the three patterns that holds across every cluster.

The per-cluster pattern is more consistent with context-level than with child-level drivers of the HL/TH signal. A subset of unusually included or unusually quiet children would produce a large uniform HL/TH gap with comparatively flat per-cluster contrasts; the observed pattern is the inverse. The pooled-across-clusters HL/TH contrast is small (§6.4) not because between-child variation masks the signal, but because the signal is context-specific. It averages to approximately zero over contexts for peer co-presence and is attenuated but not eliminated for vocal participation. The stability of the affiliation asymmetry across contexts further indicates that the pattern is not driven by a subset of children.

One caveat concerns statistical power. With 13 children and no omnibus test of the interaction block, the evidence for context-dependence is based on the pattern of per-cluster contrasts rather than on a

single summary p -value. Within that pattern, several per-cluster contrasts reach $p < 0.01$ (and some < 0.001), while the pooled child-level HL/TH contrast is small for peer co-presence and moderate for vocal participation, and not uniformly non-zero across clusters.

7.2. Methodological Implications

Four design choices were consequential for the analysis.

The first concerns context-conditioning. The unconditional HL/TH comparison in peer co-presence is small (HL $\sim 39\%$ vs. TH $\sim 40\%$, a -1.1 pp difference) and, reported in isolation, would support an inference of no detectable group difference at conventional thresholds. Context-conditioning identifies three larger context-specific deviations of opposing sign that attenuate one another at the pooled level: HL above TH in Cluster 2 ($+7.8$ pp), and HL below TH in Clusters 1 (-4.3 pp) and 4 (-5.3 pp). The unconditional vocal-participation gap is moderate (HL ~ 2.4 vs. TH ~ 3.0 utterances per minute) and runs in the same direction as the within-context effects. It averages over three sources: the two contexts where the gap is large and significant (Clusters 1 and 3), the dispersed-transition context where the gap is sizeable but only marginal after correction (Cluster 0), and the three contexts (Clusters 2, 4, and 5) where the effect is smaller and not significant. Unsupervised context discovery renders context-specific effects in either direction observable and separates the statistical instability of the adult-vocally-dominated contexts from their effect size.

The second concerns the role of the random intercept. The per-child random intercept in the LMM affects interpretation as well as estimation. It absorbs stable between-child variation in the outcome before the fixed effects are estimated, so the HL/TH signal is identified independently of that individual variation. A simpler ANOVA would report a non-significant main effect of hearing status on vocal participation and would support the interpretation that no gap is present. The LMM clarifies the meaning of that null: the hearing-status signal is concentrated in the context interactions rather than in a single overall difference. These are distinct statements, and only the second is consistent with the per-cluster pattern above.

The third concerns the multivariate treatment of inclusion. The three markers were analysed separately rather than combined into a single composite score. A composite score averaging time-in-group, vocalisation rate, and affiliation composition would have obscured the main findings, since the three markers track the demand and social structure of the surrounding context along distinct dimensions. A single child may, for example, exhibit physical co-presence with low vocal output in a noisy peer group, vocal participation in a quiet small-group setting, and grouped time disproportionately spent with mixed-diagnosis peers. The three markers jointly capture information along three axes that a univariate score cannot represent.

The fourth concerns how directly each marker operationalises the construct introduced in §5.6. The empirical pattern tracks this directness: affiliation, the closest proxy for cross-group contact, reaches significance in four of six contexts; vocal participation in two, with one further marginal; co-presence in three. Response-contingency, peer-nomination, or dyadic-affiliation analyses would address the underlying construct more directly.

7.3. Limitations

Several limitations qualify the claims above.

Single setting and small N. The pipeline is evaluated on a single cohort of 13 children. Both the shape of the six contexts and the directions of the HL/TH differences may be specific to this classroom,

age group, and curriculum. Multi-site replication is required before these findings can be treated as general. The Cluster 2 HL-above-TH peer-co-presence effect ($d = +1.59$), the Cluster 1 HL-below-TH peer-co-presence effect ($d = -1.84$), and the Cluster 4 HL-below-TH peer-co-presence effect ($d = -1.65$) each draw on under 10% of the observed day and should be regarded as the least robust of the reported effects.

Multiple comparisons. Significance in the main analysis is judged on a Benjamini–Hochberg FDR correction applied within each marker family, the six within-context contrasts of one outcome (§5.7); the 6 exposure tests serve as a balance check rather than hypotheses and are reported uncorrected. A strict family-wise correction (such as Bonferroni) over all 36 tests was judged inappropriate for an exploratory study with 13 children, since the resulting conservatism would substantially reduce statistical power. Of the 17 per-cluster contrasts that pass FDR, 14 also survive a within-family Bonferroni correction ($\alpha = 0.05/6 \approx 0.00833$): all eight same-diagnosis and mixed-diagnosis affiliation contrasts (raw $p < 0.001$ in every case), the Cluster 2 and Cluster 4 co-presence effects, the Cluster 1 vocal-participation effect, and three of the four homophily effects. Three contrasts near the threshold fail Bonferroni: the Cluster 3 vocal-participation effect ($p = 0.016$), the Cluster 1 co-presence effect ($p = 0.010$), and the Cluster 4 homophily effect ($p = 0.022$). The reported findings are therefore mostly but not uniformly robust to a stricter family-wise correction, with these three near-threshold effects the most sensitive to the choice of correction method.

Data-driven cluster count. $K = 6$ is selected by a stability-aware BIC rule over $K = 2\text{--}20$ (§5.5.4): among K values whose label assignments reproduce across random initialisations (mean pairwise ARI ≥ 0.80), the one with the lowest mean BIC is chosen. The recovered K is therefore not the unconditional argmin of BIC but the most parsimonious choice that combines low BIC with reproducibility. The slightly lower mean BIC at $K = 7$ ($\Delta\text{BIC} \approx 11$) is not used because the seven-cluster partition does not reproduce across seeds (mean pairwise ARI 0.69), and the gap is small on the Kass–Raftery scale of BIC differences [46]. The six-cluster structure is close to, but not identical to, the eight categories from the independent manual coding, which is a rough comparison rather than independent validation.

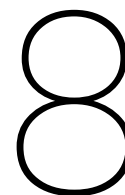
Orientation is shoulder rather than gaze. As described in Chapter 4, the orientation used in the spatial-group affinity is a shoulder (body) orientation derived from the child’s two body-worn UWB tags, rather than head direction or gaze. In young children, shoulder orientation can diverge from the direction of visual attention [41]. Detectors of conversational groups typically employ a cone of attention rather than a single orientation vector [39]. The combination of a single vector and a shoulder reference frame entails that the peer-co-presence measure carries an unquantified amount of both false-positive and false-negative group membership. For this reason, the group detector of Tian et al. [9] uses proximity alone, with orientation entering only their separate pairwise social-contact measure. This study’s affinity instead retains a mutual-orientation term (§5.2.3), and its inclusion therefore constitutes a deliberate but untested departure from that prior work. Because the spatial-feature table could not be recomputed within the scope of this thesis, the effect of the orientation term on the detected groups was not evaluated, and the peer-co-presence and affiliation results should be interpreted with that caveat.

Weak circularity in cluster labels. A mild form of circularity is present in the design: a child’s own behaviour exerts a small influence on the context label subsequently used to analyse that child. Cluster labels are shared across all children in a minute, and each child contributes $1/|C_t|$ of the room average that determines that minute’s label: as low as $\approx 8\%$ when all 13 children are present, and approximately 10% on average given a mean attendance of about ten children per jointly observed minute. A child’s own behaviour therefore has a modest but non-zero influence on the label of the context in which their HL/TH comparison is made. The effect is attenuated by averaging across the children present but not

eliminated, producing a weak coupling between context discovery and comparison.

Behavioural markers are proxies, not inclusion. The pipeline does not observe whether initiations are taken up, whether peers seek the child out, or whether the child is treated as central or peripheral, so the context-specific patterns reported here are best read as exclusion-risk markers rather than as direct measures of exclusion. Even affiliation, the closest of the three markers to the underlying construct, indexes cross-group contact rather than mutual preference, which would require peer-nomination or observer-coded data.

Baseline-rate asymmetry in peer affiliation. The homophily-index analysis is partially confounded by cohort composition. As derived in §6.4.3, cohort composition alone gives the TH homophily index a baseline advantage of approximately 0.08 over the HL index under purely random dyadic affiliation. The observed HL/TH gap across the four significant clusters is 0.12–0.19, about 0.04–0.11 above this baseline, so baseline availability accounts for a substantial share of the asymmetry and preference for the remainder. A null-model comparison (e.g. random re-assignment of diagnosis labels within each minute, matching on observed spatial group sizes) would quantify baseline’s contribution more precisely than this rough subtraction.



Societal Impact

This study develops a method for measuring, from wearable sensors, when and for whom inclusion breaks down in an inclusive preschool. The objective carries both potential benefits and substantive risks.

The benefits concern the resolution and coverage of the measurement. The approach can render observable patterns that are difficult to detect through manual observation: the classroom activities in which children with hearing loss are most peripheral, and those that support participation. Passive continuous sensing covers a full day at fine temporal resolution, which manual observation cannot achieve, and does so without directing observer attention to individual children. Applied appropriately, the resulting information could inform the structuring of classroom activities to support broader participation and could support earlier and better-targeted intervention.

The risks concern the same properties of the technology. Continuous monitoring of young children who cannot themselves consent requires informed consent from parents and guardians and oversight by the institution. Even under de-identified coding, fine-grained position and audio data carry a re-identification risk and must be stored under strict access control. The behavioural markers used here are proxies for inclusion (§7.3); treating them as direct measures could result in a child being mislabelled as excluded or in the persistent labelling of a child with hearing loss. The outputs should therefore be regarded as signals that warrant further human examination, not as automated judgements about a child. Finally, a tool that renders children's social difficulties observable carries a misuse risk for surveillance or ranking rather than support; its application should remain aligned with the objective of improving inclusion, with educators and families involved in the interpretation and use of any findings.

9

Conclusion

This study adapted a multimodal sensing pipeline to investigate social inclusion among children with hearing loss in an inclusive preschool. It combined UWB tracking with LENA audio, discovered latent activity contexts with a GMM, and fitted an LMM with a per-child random intercept to compare HL and TH children on three sensor-derivable behavioural markers of inclusion within each context.

Six activity contexts were recovered, each with a distinct sensor profile and given a descriptive label corresponding to a recognisable phase of the inclusive preschool day. The cluster count was selected by a stability-aware BIC rule over $K = 2\text{--}20$ that combines lowest mean BIC with a partition-reproducibility requirement across random initialisations (§5.5.4); the resulting $K = 6$ is of similar magnitude to, but not identical to, the eight activity types identified by an independent manual coding of the routine, which served only as an informal qualitative comparison rather than a ground truth, with no quantitative correspondence analysis performed.

Within these contexts, peer co-presence differs between HL and TH in three of six contexts: an HL-above-TH effect in independent / parallel work (+7.8% of the minute, $q < 0.0001$, $d = +1.59$; the compensatory-presence account is one of several plausible interpretations, alongside teacher-directed seating and activity-area composition), a HL-below-TH effect in seated guided work (−5.3% of the minute, $q = 0.002$, the largest HL-below-TH gap), and a HL-below-TH effect in peer-driven activity (−4.3% of the minute, $q = 0.020$). Vocal participation shows an HL-below-TH gap in the contexts that combine high peer communicative demand with reduced adult scaffolding (peer-driven activity $q = 0.034$ and adult-scaffolded peer activity $q = 0.048$, both significant after FDR correction; dispersed transition in the same direction but marginal, $q = 0.071$; all $d \approx 0.8\text{--}1.0$), and in the same direction but non-significantly in the three contexts where very low auditory overlap or adult voice dominance slows the exchange. Peer affiliation is the most consistent asymmetry, with the HL homophily index below the TH index in all six contexts (significantly in four; the two non-significant gaps are within the magnitude expected from the 6:7 cohort composition alone, while the four significant gaps exceed that baseline) and HL children spending more grouped time in mixed spatial groups than TH children across all six (significantly in four).

Context-conditioning identifies effects that a whole-day comparison averages out. All of the above rests on a single classroom; if multi-site replication did not reproduce the six-context structure, the activity-type labels in §6.1 would need to be treated as classroom-specific rather than generic to inclusive

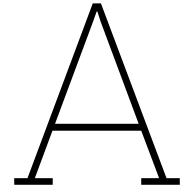
preschool practice. The contribution of this study is less in the specific numerical estimates than in indicating that the underlying questions are addressable through passive sensing in principle, pending replication: the environmental demands under which exclusion occurs, the children to whom it applies, and the peers from whom it arises.

10

Future Work

Several extensions follow from the limitations identified in §7.3. Multi-site replication would establish whether the six contexts and the HL/TH directions reported in §6.4 (concentrated in Clusters 1 through 5, with Cluster 0 the only context in which no marker shows a significant gap) are specific to this classroom or generic features of inclusive preschool practice. A longitudinal follow-up over an academic year would separate stable from transient individual differences and provide a larger sample against which to estimate the random-intercept structure. The choice of $K = 6$ is the stability-aware BIC selection on this classroom's data; deriving a candidate activity-type set from an independently coded cohort would test whether the six-context structure generalises beyond this setting.

Replacing the single-vector orientation model with a cone-of-attention model and comparing the resulting spatial group memberships against human-coded ground truth would quantify the measurement error currently embedded in peer co-presence. Coupling context discovery with finer-grained turn-level analysis (including response contingency and initiation success, neither captured at the one-minute epoch) would extend the pipeline toward the proximal exclusion mechanisms identified in the background literature and reduce the gap between the behavioural markers and the inclusion construct they proxy.



Software and Reproducibility

A.1. Software

The analysis pipeline is implemented in Python (3.10+) using standard scientific libraries. The libraries used are pandas and NumPy for data handling, SciPy for non-parametric tests, scikit-learn for the Yeo–Johnson power transformation and the GMM, and statsmodels for the LMM (fit by restricted maximum likelihood). Figures are produced with matplotlib and seaborn. The Dominant Sets spatial group extractor is adopted unchanged from the reference pipeline [10]. The specific package versions used for the numerical results reported here are pinned in the released code repository.

A.2. Reproducibility

The GMM is fit with a fixed random seed and ten random restarts at each K in the BIC sweep, so K -selection and clustering reproduce exactly on the same input. The Dominant Sets iteration inside the spatial group extractor starts from the uniform point of the simplex (every participant weighted equally; the standard initialisation of Pavan and Pelillo [40]) and runs to a fixed convergence tolerance (10^{-6}) under a fixed iteration cap, so it has no random component and the minute-level spatial-feature table reproduces exactly across independent runs. The entire pipeline is therefore deterministic on the same input, and all numerical results in Chapter 6 reproduce exactly on re-running. The statistical analysis is restricted to minutes marked present in the daily attendance record, and the same filter is applied upstream of both the results figures and the LMM tables, so Figures 6.2–6.4 and Tables 6.4–6.8 are computed on the same 21,847-minute subset.

A.3. AI-Tool Disclosure

Generative AI tools (large language models) were used during this thesis. Following the TU Delft guidelines on the use of AI, their use is disclosed by stage below.

- **Idea generation and research design.** No generative AI tool was used to fully formulate the research questions, hypotheses, or study design, these were developed by the author in discussion with the supervisors.

- **Understanding theory and techniques.** Large language model assistance was used as a supplementary aid to clarify statistical and computational concepts (for example, the LMM, the Benjamini–Hochberg correction, and the Dominant Sets formulation). All such explanations were checked against the cited primary sources, which remain the basis for the methods described here.
- **Literature search and reading.** The literature search was carried out by the author through database tools. Generative AI was used for organising and formatting references. All cited works were read and verified by the author.
- **Code.** Large language model assistance was used during development for drafting and debugging code. All analysis logic was specified, reviewed, and tested by the author.
- **Writing and editing.** Large language model assistance was used to revise and restructure prose.

No generative AI tool was used to generate, run, or interpret the statistical analyses, or to produce any numerical result; these were computed by the analysis pipeline described in this thesis.

References

- [1] Y. H. Xie, M. Potměšil, and B. Peters. “Children Who Are Deaf or Hard of Hearing in Inclusive Educational Settings: A Literature Review on Interactions With Peers”. In: *Journal of Deaf Studies and Deaf Education* 19.4 (2014), pp. 423–437. doi: 10.1093/deafed/enu017.
- [2] S. D. Antia, K. H. Kreimeyer, and K. K. Metz. “Peer interactions of deaf and hard-of-hearing children”. In: *The Oxford handbook of deaf studies, language, and education*. Ed. by M. Marschark and P. E. Spencer. 2nd. Vol. 1. New York: Oxford University Press, 2011, pp. 195–207.
- [3] Deborah Mood et al. “Relational Factors in Pragmatic Skill Development: Deaf and Hard of Hearing Infants and Toddlers”. In: *Pediatrics* 146.Suppl. 3 (2020), S246–S261. doi: 10.1542/peds.2020-0242D.
- [4] Juliene Madureira Ferreira. “An Embodied View on Collaboration in Early Childhood Education: Combining Microanalysis and Introspective Analysis of Experiences to Understand Meaning-Making Between Children With and Without Intellectual Disabilities”. In: *Human Arenas* (2023). doi: 10.1007/s42087-023-00380-4.
- [5] Patrick J. Brice and Gretchen Strauss. “Deaf adolescents in a hearing world: a review of factors affecting psychosocial adaptation”. In: *Adolescent Health, Medicine and Therapeutics* 7 (2016), pp. 67–76. doi: 10.2147/AHMT.S60261.
- [6] S. M. Murphy, D. M. Faulkner, and L. R. Farley. “The Behaviour of Young Children with Social Communication Disorders During Dyadic Interaction with Peers”. In: *Journal of Abnormal Child Psychology* 42.2 (2014), pp. 277–289. doi: 10.1007/s10802-013-9772-6.
- [7] Batya Elbaum, Lynn K. Perry, and Daniel S. Messinger. “Investigating children’s interactions in preschool classrooms: An overview of research using automated sensing technologies”. In: *Early Childhood Research Quarterly* 66 (2024), pp. 147–156. doi: 10.1016/j.ecresq.2023.10.005.
- [8] D. W. Irvin et al. “Capturing Talk and Proximity in the Classroom: Advances in Measuring Features of Young Children’s Friendships”. In: *Early Childhood Research Quarterly* 57 (2021), pp. 102–109. doi: 10.1016/j.ecresq.2021.05.003.
- [9] TUDelft-SPC-Lab. ICDL2025: Code related to the ICDL 2025 publication “Groups Matter: Investigating the Effects of Homophily in Child Interactions in an Inclusive Classroom”. <https://github.com/TUDelft-SPC-Lab/ICDL2025>. 2025.
- [10] TUDelft-SPC-Lab. *Group-detection*. <https://github.com/TUDelft-SPC-Lab/group-detection>. Apr. 2022.
- [11] Stefanie Krijger et al. “Listening Difficulties of Children With Cochlear Implants in Mainstream Secondary Education”. In: *Ear and Hearing* 41.5 (2020), pp. 1172–1186. doi: 10.1097/AUD.0000000000000835.
- [12] Brenda M. S. Da Silva et al. “Being Deaf in Mainstream Schools: The Effect of a Hearing Loss in Children’s Playground Behaviors”. In: *Children* 9.7 (2022), p. 1091. doi: 10.3390/children9071091.

- [13] J. Bobzien et al. "An Observational Study of Social Communication Skills in Eight Preschoolers with and Without Hearing Loss During Cooperative Play". In: *Early Childhood Education Journal* 41 (2013), pp. 339–346. doi: 10.1007/s10643-012-0561-6.
- [14] Yuan Tian et al. "Groups Matter: Investigating the Effects of Homophily in Child Interactions in an Inclusive Classroom". In: *2025 IEEE International Conference on Development and Learning (ICDL)*. Prague, Czech Republic: IEEE, Sept. 2025, pp. 1–6. ISBN: 979-8-3315-4343-3.
- [15] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27.1 (2001), pp. 415–444. doi: 10.1146/annurev.soc.27.1.415.
- [16] Harriet R. Tenenbaum et al. "How Effective is Peer Interaction in Facilitating Learning? A Meta-Analysis". In: *Journal of Educational Psychology* 112.7 (2020), pp. 1303–1319. doi: 10.1037/edu0000436.
- [17] Joshua J. Guyer et al. "Paralinguistic Features Communicated through Voice can Affect Appraisals of Confidence and Evaluative Judgments". In: *Journal of Nonverbal Behavior* 45.4 (2021), pp. 479–504. doi: 10.1007/s10919-021-00374-2.
- [18] Benjamin Waber et al. "A voice is worth a thousand words: The implications of the micro-coding of social signals in speech for trust research". In: *Handbook of Research Methods on Trust*. Ed. by Fergus Lyon, Guido Möllering, and Mark N. K. Saunders. Edward Elgar Publishing, 2012. Chap. 23, pp. 249–258.
- [19] Rainer Banse and Klaus R. Scherer. "Acoustic profiles in vocal emotion expression". In: *Journal of Personality and Social Psychology* 70.3 (1996), pp. 614–636. doi: 10.1037/0022-3514.70.3.614.
- [20] David A. Puts, Steven J. C. Gaulin, and Katherine Verdolini. "Dominance and the evolution of sexual dimorphism in human voice pitch". In: *Evolution and Human Behavior* 27.4 (2006), pp. 283–296. doi: 10.1016/j.evolhumbehav.2005.11.003.
- [21] Henry S. Cheang and Marc D. Pell. "The sound of sarcasm". In: *Speech Communication* 50.5 (2008), pp. 366–381. doi: 10.1016/j.specom.2007.11.003.
- [22] Xin Luo, Qian-Jie Fu, and John J. III Galvin. "Vocal Emotion Recognition with Cochlear Implants". In: *INTERSPEECH 2006 – ICSLP*. 2006, pp. 1830–1833. doi: 10.21437/Interspeech.2006-505.
- [23] Tracey Silveira-Zaldivar, Gül Özerk, and Kamil Özerk. "Developing Social Skills and Social Competence in Children with Autism". In: *International Electronic Journal of Elementary Education* 13.3 (2021), pp. 341–363. doi: 10.26822/iejee.2021.195.
- [24] Chris Forlin. "Inclusive education in Australia ten years after Salamanca". In: *European Journal of Psychology of Education* 21.3 (2006), pp. 265–277. doi: 10.1007/BF03173415.
- [25] Michaela Socher et al. "Pragmatic Language Skills: A Comparison of Children With Cochlear Implants and Children Without Hearing Loss". In: *Frontiers in Psychology* 10 (2019), p. 2243. doi: 10.3389/fpsyg.2019.02243.
- [26] R. M. Fasano et al. "A Granular Perspective on Inclusion: Objectively Measured Interactions of Preschoolers with and Without Autism". In: *Autism Research* 14.8 (2021), pp. 1658–1669. doi: 10.1002/aur.2526.
- [27] T.-J. Lin et al. "Peer Interactions in Preschool Inclusive Classrooms: The Roles of Pragmatic Language and Self-Regulation". In: *Exceptional Children* 85.4 (2019), pp. 432–452. doi: 10.1177/0014402919828364.

- [28] David Wood, Jerome S. Bruner, and Gail Ross. "The role of tutoring in problem solving". In: *Journal of Child Psychology and Psychiatry* 17.2 (1976), pp. 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x.
- [29] Lill-Johanne Eilertsen et al. "Collaborative, responsive behavior in peer interaction: Children with and without Down syndrome and hearing impairment". In: *Journal of Interactional Research in Communication Disorders* 15.2 (2024), pp. 1–27. doi: 10.1558/jircd.24422.
- [30] Frank Iglehart. "Speech Perception in Classroom Acoustics by Children With Hearing Loss and Wearing Hearing Aids". In: *American Journal of Audiology* 29.1 (2019), pp. 6–17. doi: 10.1044/2019_AJA-19-0010.
- [31] Adva Eichengreen et al. "Using wearable sensors to explore schoolyard interactions of mainstreamed deaf and hard-of-hearing preadolescents". In: *Social Development* (2024), e12755. doi: 10.1111/sode.12755.
- [32] Mary Ford et al. *The LENA Language Environment Analysis System: Audio Specifications of the DLP-0121*. Tech. rep. LENA Research Foundation, 2008.
- [33] Jill Galkerson and Jeffrey A. Richards. *The LENA natural language study*. Technical Report LTR-02-2. LENA Foundation, 2008.
- [34] Dongxin Xu et al. *The LENA Language Environment Analysis System: The Interpreted Time Segments (ITS) File*. Technical Report LTR-04-2. Boulder, CO: LENA Foundation, 2009.
- [35] Alejandrina Cristia et al. "A thorough evaluation of the Language Environment Analysis (LENA) system". In: *Behavior Research Methods* 53.2 (2020), pp. 467–486. doi: 10.3758/s13428-020-01393-5.
- [36] Ye Wang et al. "A Systematic Review of the Use of LENA Technology". In: *American Annals of the Deaf* 162.3 (2017), pp. 295–311. doi: 10.1353/aad.2017.0028.
- [37] Adam Kendon. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge, UK: Cambridge University Press, 1990.
- [38] Hayley Hung and Ben Kröse. "Detecting F-formations as Dominant Sets". In: *Proceedings of the 13th International Conference on Multimodal Interaction (ICMI)*. ACM, 2011, pp. 231–238. doi: 10.1145/2070481.2070525.
- [39] Francesco Setti et al. "F-Formation Detection: Individuating Free-Standing Conversational Groups in Images". In: *PLOS ONE* 10.5 (2015), e0123783. doi: 10.1371/journal.pone.0123783.
- [40] Massimiliano Pavan and Marcello Pelillo. "Dominant sets and pairwise clustering". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.1 (2007), pp. 167–172. doi: 10.1109/TPAMI.2007.250608.
- [41] Hanako Yoshida and Linda B. Smith. "What's in view for toddlers? Using a head camera to study visual experience". In: *Infancy* 13.3 (2008), pp. 229–248. doi: 10.1080/15250000802004437.
- [42] Khalid N. Alasim. "Participation and Interaction of Deaf and Hard-of-Hearing Students in Inclusion Classroom". In: *International Journal of Special Education* 33.2 (2018), pp. 493–506.
- [43] Femke Rademaker et al. "Applying the Contact Theory in Inclusive Education: A Systematic Review on the Impact of Contact and Information on the Social Participation of Students With Disabilities". In: *Frontiers in Education* 5 (2020), p. 602414. doi: 10.3389/educ.2020.602414.
- [44] In-Kwon Yeo and Richard A. Johnson. "A new family of power transformations to improve normality or symmetry". In: *Biometrika* 87.4 (2000), pp. 954–959. doi: 10.1093/biomet/87.4.954.

- [45] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [46] Robert E. Kass and Adrian E. Raftery. "Bayes Factors". In: *Journal of the American Statistical Association* 90.430 (1995), pp. 773–795. DOI: 10.1080/01621459.1995.10476572.
- [47] Geoffrey J. McLachlan and David Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. New York: Wiley, 2000. ISBN: 978-0-471-00626-8. DOI: 10.1002/0471721182.
- [48] Douglas Bates et al. "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1 (2015), pp. 1–48. DOI: 10.18637/jss.v067.i01.
- [49] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. 2nd. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.