Available Online: 3 November 2026 DOI: 10.54254/3049-7248/2026.38658

Teachers' prioritisation of quiet and disruptive student disengagement in a mouse tracking study

Beiyi Cai

University College London, London, UK

changgeyeer@gmail.com

Abstract. Teachers must decide which students need support when several classroom demands compete for a response. This study examined support prioritisation when quiet and disruptive disengagement were explicitly presented. Forty practising teachers completed an online two-alternative forced-choice mouse-tracking task with 32 formal trials across quiet and noisy classroom environments. No statistically reliable overall disruptive-choice advantage was detected, and the noisy environment did not significantly increase disruptive selections. Adjusted models of log response time indicated shorter total response time and post-initiation intervals in the noisy condition; unadjusted participant-level mean-difference intervals included zero. Reconstructed maximum deviation and absolute net trajectory area showed no reliable environment or final-selection effects. An exploratory pooled drift-diffusion model described a starting point slightly closer to the disruptive boundary and drift toward the quiet boundary, but participant bootstrap uncertainty and fit diagnostics limited mechanistic interpretation. No environment-specific diffusion parameter survived adjustment across four comparisons. The findings concern explicit support choices between two displayed profiles. They do not establish equivalent preferences or describe which students teachers would spontaneously notice in a live classroom.

Keywords: teacher noticing, student disengagement, support prioritisation, mouse tracking, classroom environment, drift-diffusion model, professional vision

1. Introduction

Classroom teaching requires teachers to monitor multiple students while maintaining instructional flow, managing peer interaction, and responding to emerging behavioural and learning needs. These demands make exhaustive processing of all available information unrealistic. Bounded rationality describes how decision makers operate under limits of time and cognitive capacity, relying on selective processing rather than complete analysis [1]. In teaching, this selectivity is functional: teachers must decide which cues are relevant enough to interpret and act on. Contemporary models of teacher competence similarly distinguish perception, interpretation, and decision making as related but separable situation-specific processes linking knowledge to classroom action [2, 3].

A conspicuous behaviour can attract attention without being judged the highest-priority support need. Research on teacher noticing therefore needs to distinguish the information available to a teacher from the

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

interpretation and response that follow. Reviews show that instruments capture different components of noticing and professional vision [4, 5]. Eye-tracking evidence also identifies expertise-related differences in visual allocation: experienced teachers tend to distribute gaze more evenly across students [6]. These findings establish the importance of specifying the stage being measured; they do not, by themselves, predict whom teachers will choose to support.

1.1. From noticing to support prioritisation

Professional vision includes interpreting classroom events and considering possible responses. In video-based classroom-management tasks, experienced teachers interpreted more information and generated more alternative responses than novices [7]. Support prioritisation adds a specific decision to this sequence: when more than one student shows difficulty, whose need should be addressed first? This decision concerns the meaning and urgency assigned to available cues, rather than simply whether a cue attracts visual attention.

This question matters because teacher support is associated with students' engagement in school. A meta-analysis of 143 studies reported in 141 articles found a moderate positive association between teacher support and school engagement [8]. Here, school engagement refers to students' engagement with learning and school, not an attribute of the school as an institution. Although this association does not establish the effects of a particular support decision, it provides an educational reason to examine how teachers allocate support when needs compete.

1.2. Disruptive and quiet disengagement

Student engagement encompasses behavioural, emotional, and cognitive involvement in learning [9, 10]. Reduced involvement can be expressed through overt disruption or through withdrawal. For the comparison developed here, disruptive disengagement denotes behaviours that interrupt classroom activity, such as making noise or distracting peers; quiet disengagement denotes less overt behaviours, such as reduced participation or passive inattention. These labels describe the behavioural contrast used in the task, rather than an exhaustive classification of disengagement or independently validated levels of perceptual salience.

Attentional-selection theory distinguishes stimulus-driven salience from selection shaped by goals and prior experience [11]. This distinction matters for support decisions because the cue that attracts attention need not be the cue assigned the greatest educational urgency. Overt disruption directly interrupts classroom activity and could receive priority if this immediate interference dominates teachers' decisions. Teacher-professional-vision research shows that classroom information is selectively organised rather than processed uniformly [3, 12], but does not establish that disruption must dominate support choices. H1 tests that specific prediction.

H1. Teachers will prioritise the student displaying disruptive disengagement more frequently than the student displaying quiet disengagement.

A competing interpretation follows from the decision goal. When the question is who should receive support first, withdrawal and low participation can signal a need that would otherwise receive little attention. This support-oriented interpretation could weaken or reverse a disruptive-choice advantage. H1 is therefore a test of the disruptive-priority prediction, not a prediction derived from the competing support-goal account. Because the task did not compare different decision prompts, an unsupported H1 would not establish that the support wording caused teachers' choices.

1.3. Classroom environment and prioritisation

Environmental conditions provide a second possible influence on support decisions. Reviews of noise and cognition report effects that vary with task demands and individual differences, including null as well as

adverse effects [13, 14]. Thus, a noisy classroom condition need not change every aspect of responding in the same way. It could affect which student is selected, response speed, or both.

If competing classroom input increases reliance on immediately disruptive cues, the disruptive-priority account predicts more disruptive selections in the noisy condition. This is a hypothesis about the classroom-environment contrast used here, not an established consequence of auditory noise. In the task, the visual background and accompanying audio changed together; the conditions therefore combined visual and auditory information. H2 tests whether that combined contrast shifts support priority toward disruptive disengagement.

H2. The likelihood of prioritising the student displaying disruptive disengagement will be higher in the noisy classroom condition than in the quiet classroom condition.

1.4. Process tracing and the present study

Final choices identify the selected option but provide limited information about response development. Mouse tracking records movement continuously during a choice, allowing response timing and trajectory geometry to be analysed alongside the final selection [15-17]. These measures are sensitive to task implementation and require care when relating movement to cognitive processes. They complement a prioritisation outcome but do not directly measure where a teacher looked.

Exploratory drift-diffusion modelling provides a complementary account of choice and response time through drift rate (v), boundary separation (a), starting point (z), and non-decision time (t_0) [18]. These parameters describe accumulation direction, response caution, initial asymmetry, and time outside accumulation, respectively. The task specified no correct response, so the model was interpreted in terms of preference-related accumulation. Its parameters were estimated at the aggregate level and were not treated as direct measures of visual attention.

The present study displayed quiet and disruptive student profiles simultaneously and asked teachers which student they would support first. H1 concerned an overall disruptive-choice advantage, and H2 concerned a stronger disruptive preference in the noisy classroom condition. Secondary analyses examined response timing and trajectories; diffusion modelling and participant-group comparisons were exploratory. By making both profiles available, the design tested explicit support prioritisation rather than spontaneous detection of disengagement.

2. Method

2.1. Participants and data screening

Participants were recruited through convenience and snowball sampling in online teacher communities and professional education networks. Recruitment materials described the study as an online classroom decision-making task for practising teachers. Participation was voluntary. The task presented an electronic consent screen before the instructions and allowed participants to indicate consent by clicking an agreement button. Task files used participant numbers rather than names. Study procedures were approved before data collection.

Eligible participants were practising teachers who could complete the task on a desktop or laptop computer. Because the study required continuous cursor recording, touchscreen responses were not permitted. Participants were excluded if they did not complete the formal task or failed to provide usable choice or trajectory data. The final analytic sample comprised 40 teachers. Each participant contributed 32 valid formal trials, yielding 1,280 valid final decisions. Two practice trials per participant were excluded. Twenty-eight formal-trial attempts exceeded the 12-second response limit and were excluded; those trials were repeated

until a valid response was obtained, so all retained participants contributed one valid response to each formal trial.

Sample-size planning used a two-tailed paired-samples t-test benchmark with $\alpha = 0.05$, *target power* = 0.80, and an assumed medium standardised effect $d_z = 0.50$. These G***Power-style* settings require a minimum of 34 participants; a noncentral-t recalculation *gives power* = 0.808 at $N = 34$ and 0.795 at $N = 33$. Forty participants were recruited to allow for unusable recordings. The assumed effect size was not estimated from the observed study effect. This benchmark was not a model-specific power calculation for the binary mixed-effects outcome or a post hoc demonstration of power for null results. Participant characteristics are reported in Table 1.

Table 1. Participant characteristics

Characteristic	*n* (%) or *M* (*SD*)
Gender	
Women	26 (65.0%)
Men	14 (35.0%)
Age category	
20-30 years	23 (57.5%)
30-40 years	15 (37.5%)
40-50 years	2 (5.0%)
Highest qualification	
Bachelor's degree	20 (50.0%)
Master's degree	19 (47.5%)
Doctoral degree	1 (2.5%)
Teaching experience (years)	4.93 (4.43)
Form-teacher experience (years)	2.10 (3.52)
No form-teacher experience	19 (47.5%)
Primary subject area	
Humanities	20 (50.0%)
Sciences	18 (45.0%)
Arts	1 (2.5%)
Other	1 (2.5%)
Educational level taught	
Junior secondary	29 (72.5%)
Senior secondary	8 (20.0%)
University	3 (7.5%)
Classroom-noise item score	2.78 (1.03)

Note: Values are frequencies with percentages unless otherwise indicated. *M* = Mean; *SD* = Standard Deviation.

2.2. Design

The study used an online within-subject two-alternative forced-choice mouse-tracking design. On every formal trial, one student profile represented disruptive disengagement and the other represented quiet disengagement. Participants selected which student they would prioritise for immediate classroom intervention or support.

The 32 formal trials formed a fully crossed combination of eight classroom situations, two classroom-environment conditions (quiet and noisy), and two stimulus-gender versions (male and female). Each situation therefore appeared once in every environment-by-stimulus-gender combination. The left-right position of the disruptive profile was randomised independently on each trial, and trial order was randomised for each participant. Figure 1 presents representative materials and the trial procedure.

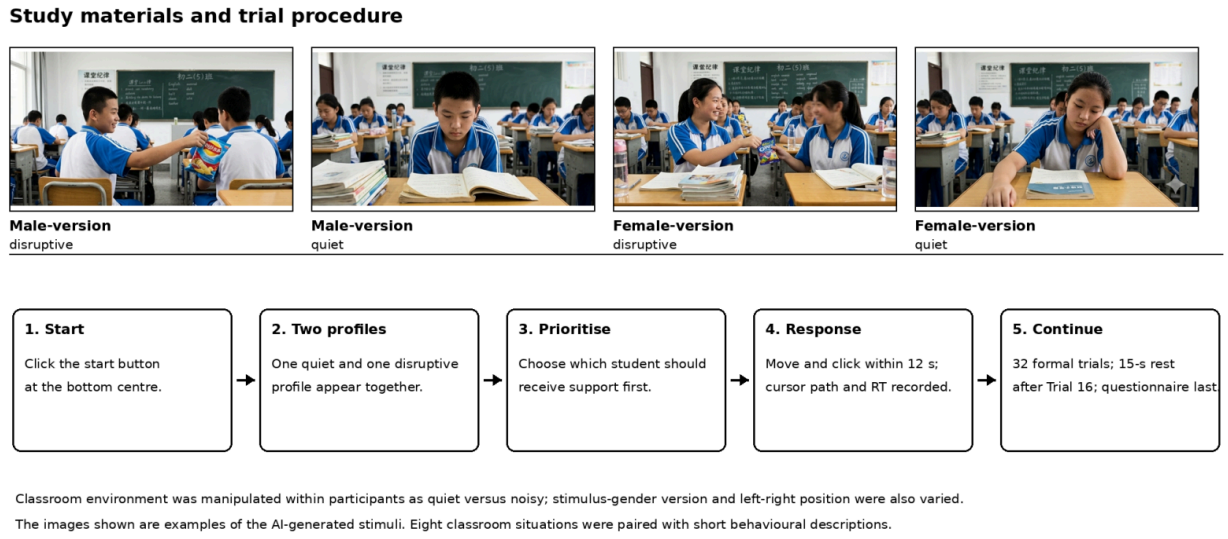

Figure 1. Study materials and trial procedure

Note: The upper panels show examples from the AI-generated stimulus set; each image was paired with a behavioural description. The lower panels show the trial sequence. The schematic does not reproduce the exact web interface. The formal task crossed eight situations, two classroom environments, and two stimulus-gender versions.

2.3. Materials

Eight paired classroom scenarios contrasted disruptive and quiet disengagement. The implemented English descriptions are reproduced in Appendix B. Disruptive descriptions included making noise, distracting peers, or resisting participation; quiet descriptions included staring blankly, resting the head on the desk, or withdrawing from an activity. Descriptions contained 12–18 whitespace-delimited words and were not exactly matched in length. Identical descriptions were used across the two stimulus-gender versions and the two environment conditions. The task did not specify a correct choice.

Two practising secondary-school teachers reviewed the candidate profiles and descriptions for classroom plausibility, and their feedback informed revisions. This informal review did not provide independent quantitative validation of salience, severity, or support need.

Each behavioural description was paired with an AI-generated student image developed iteratively through Gemini. The final task contained 32 student images: eight situations crossed with two stimulus-gender versions and two disengagement types. Each image was reused across the two environment conditions. The

retained conversation excerpts document requests for classroom scenes, corresponding male and female depictions, and revisions to composition. Generation records and representative images are presented in Appendix B. The images were not controlled morphs, and disengagement type and stimulus-gender version are treated as contrasts between the implemented stimulus sets.

Classroom environment combined visual and auditory information. Separate classroom background images and looping audio files were used for the quiet and noisy conditions; the quiet condition was not implemented as silence. The background images differed in scene composition, not just visual activity. Background and audio presentation began on the start screen and continued during the choice screen. Audio playback success and sound pressure level were not recorded. The manipulation therefore compared two classroom environments rather than isolating auditory noise.

2.4. Procedure

Participants read the instructions and adopted the role of a teacher at the beginning of a lesson. The instructions specified no correct response and asked participants to respond naturally and promptly. Two non-classroom practice trials introduced the response format and cursor movements; practice data were excluded.

Each formal trial began with a start button at the bottom centre of the screen. Clicking it displayed the two student profiles and the prompt "Please prioritise which student you would intervene/support first". Participants moved the cursor to the corresponding response button and clicked to respond. The choice screen had a 12-second limit, with its countdown hidden. Timed-out attempts were recorded and the same trial was repeated with the same layout until a valid response was obtained. After 16 formal trials, participants completed a compulsory 15-second rest. Demographic and professional-background information was collected through a separate questionnaire. The classroom-noise item asked how easily classroom noise caused irritation or distraction in everyday teaching. Its slider allowed scores from 0 to 5 without verbal endpoint anchors; observed responses ranged from 1 to 5. The numeric scores were retained without reverse scoring, and higher values were interpreted in the direction of greater reported ease of distraction rather than greater tolerance.

2.5. Measures and data processing

Total response time was measured from choice-screen onset to the final click. The program recorded cursor positions on mousemove events, with integer viewport pixel coordinates and elapsed timestamps in milliseconds; sampling was event-driven rather than at a fixed frequency. Movement-initiation time was the timestamp of the first recorded point more than 3 pixels from the first recorded cursor position. This reference position was not a separately recorded start-button click. The post-initiation interval was total response time minus this initiation timestamp. It included pauses and the interval from the last recorded movement to the final click, and therefore was not a measure of continuous motor execution alone.

Practice and timed-out attempts were excluded, retaining 1,280 formal trials with valid recorded trajectories. For each path, the first recorded position was subtracted from all coordinates. Horizontal displacement was divided by the absolute first-to-last horizontal displacement; upward vertical displacement was divided by the absolute first-to-last vertical displacement. Paths were reflected horizontally when necessary so that quiet selection ended at (-1, 1) and disruptive selection at (1, 1). All retained paths had nonzero horizontal displacement and upward vertical displacement. This endpoint-based scaling does not reconstruct the unrecorded start-button click or the exact response-button centres. Repeated timestamps within a rounded millisecond were reduced to the last sample for interpolation. Plotting trajectories were interpolated to 101 equally spaced points between the first and last recorded timestamps. Within each environment-by-selection cell, trial paths were first averaged within participants and then across participants with at least one

such choice. Figure 4 shows the participant-averaged trajectories obtained after this alignment and interpolation.

Maximum Absolute Deviation (MAD) was the largest absolute perpendicular distance from the straight first-to-last chord in the spatially normalised recorded path. Absolute net AUC was the absolute signed polygon area enclosed by the path and its chord; areas on opposite sides can cancel before taking the absolute value, so it is not the integral of unsigned deviation along path length. Both measures were computed from all spatially normalised recorded samples rather than the 101-point plotting approximation. Definitions were checked against absolute mousetrap MAD and AUC for representative paths. The task-exported pixel-based maximum deviation and custom path-length-weighted AUC were retained for verification but were not substituted for these normalised geometric measures. Path efficiency was straight-chord length divided by recorded path length in normalised coordinates. Early movement toward the unselected option was defined by the sign of horizontal position at 20% of the first-to-last recorded interval. Deviation timing used that same interval; horizontal direction changes were counted on the 101-point paths after removing zero increments. Appendix A provides the descriptive distribution of reconstructed MAD (Figure A1).

The drift diffusion model estimated four parameters jointly. Drift rate (v) was positive when accumulation favoured disruptive disengagement and negative when it favoured quiet disengagement. Boundary separation (a) indexed response caution. Starting point (z) was expressed as a proportion of the distance between the quiet boundary (0) and disruptive boundary (1), so $z = 0.50$ indicated no starting bias and $z > 0.50$ indicated a starting position closer to disruptive disengagement. Non-decision time (t_0) represented encoding and motor processes outside evidence accumulation. The DDM starting point is a latent evidence parameter and should not be interpreted as the physical starting position of the mouse.

2.6. Statistical analysis

All tests were two-tailed with $\alpha = 0.05$. H1 was assessed by comparing participant-level proportions of disruptive choices with 0.50 using a one-sample Wilcoxon signed-rank test with a normal approximation, tie adjustment, and no continuity correction. The observed proportions were described as departing from a normal distribution; this refers to their distribution across participants, not to invalid observations. H2 was assessed using the trial-level binomial mixed-effects model. Other task-level effects were secondary, and diffusion and participant-group analyses were exploratory.

Trial-level final selections were analysed using binomial generalised linear mixed-effects models with a logit link. Classroom environment, stimulus-gender version, disruptive-profile position, and standardised trial order were included as fixed effects, with random intercepts for participants and classroom situations. The Environment by Stimulus Gender interaction was tested before the corresponding main effects. Fixed effects are reported as log-odds coefficients, standard errors, Wald z statistics, p values, odds ratios, and 95% confidence intervals. The contribution of classroom environment was also evaluated by likelihood-ratio comparison with a reduced model.

Total response time, initiation time, and post-initiation interval were natural-log transformed. MAD and absolute net AUC were transformed as $log(1 + value)$; path efficiency was analysed on its original scale. Linear mixed-effects models included environment, final selection, their interaction, disruptive-profile side, stimulus-gender version, and standardised trial order. Random intercepts represented participants and situations. A situation intercept was removed when its variance was estimated at zero, as occurred for initiation time and AUC. Models used REML and Satterthwaite degrees of freedom. Reported environment contrasts compare noisy with quiet conditions on the transformed scale, averaging equally over the two selection categories; selection contrasts average equally over environments. These adjusted contrasts are not

identical to the environment coefficient at the quiet-selection reference level. Final selection was an observed response, so associations conditional on selection are not causal effects. Participant-specific environment slopes were examined as sensitivity analyses; residual quantile plots were inspected, and unadjusted participant-paired mean differences were additionally described using bootstrap intervals.

The exploratory pooled DDM was fitted to final choices and response times in seconds using rtdists 0.12-0 [19] in R 4.6.0. Disruptive selection was the upper boundary and quiet selection the lower boundary. Diffusion noise was fixed at $s = 1$; between-trial parameter variability and boundary-specific motor offsets were fixed at zero. Relative z was estimated and passed to the density function as absolute $z \times a$. The likelihood was conditioned on responding before the 12-second deadline, using the sum of upper- and lower-boundary probabilities by 12 seconds as its normalisation. Bounds were v from -5 to 5, a from 0.3 to 8, z from 0.05 to 0.95, and t_0 from 0.001 seconds to 0.001 seconds below the shortest retained response. A midpoint-fixed model, a common four-parameter model, and four models allowing one parameter at a time to vary by environment were optimised with 12 initialisations each. Likelihood-ratio tests, AIC, and BIC were reported. Benjamini–Hochberg adjustment covered the four environment-specific parameter comparisons. Pooled likelihood-ratio p values do not account for participant clustering and are treated as descriptive. Common-parameter uncertainty was assessed by 200 participant bootstrap resamples. Predicted choice probabilities and response-time quantiles were checked, with an untruncated common-model fit as a sensitivity analysis.

Exploratory participant-level analyses linked questionnaire serial numbers 1-40 to the corresponding participant data files. Teachers who selected disruptive disengagement on more than half of the 32 formal trials were compared with those who selected it on fewer than half; no participant selected it on exactly half. Mann-Whitney U tests used a normal approximation with tie adjustment and continuity correction for age category, qualification level, teaching experience, form-teacher experience, and classroom-noise item score. Fisher's exact tests were used for gender, subject area, and educational level taught. Benjamini-Hochberg correction covered these eight comparisons. The additional binary comparison of any form-teacher experience was reported with an unadjusted Fisher's exact p value and a cross-product odds ratio. Secondary choice predictors and temporal and trajectory analyses were not adjusted as a family of confirmatory tests and are interpreted cautiously. The four environment-specific DDM comparisons were adjusted separately as described above.

3. Results

3.1. Final selection

3.1.1. Overall prioritisation

Across 1,280 valid trials, quiet disengagement was selected 703 times (54.9%) and disruptive disengagement 577 times (45.1%). The participant-level mean proportion of disruptive selections was 0.451 (SD = 0.370; *Median* = 0.359). The Wilcoxon signed-rank test did not identify a statistically reliable departure from 0.50, W = 334.5, p = 0.309. The disruptive-priority prediction in H1 was not supported. The descriptive difference in aggregated choices does not establish a group-level quiet preference, and the nonsignificant test does not establish equivalence.

3.1.2. Classroom environment and stimulus-gender version

The Environment by Stimulus Gender interaction did not significantly improve the final-selection model, likelihood-ratio $\chi^2(1) = 0.19$, p = 0.665. The interaction was therefore not interpreted further. In the reduced main-effects model, classroom environment did not significantly predict final selection, b = -0.10, SE = 0.16, z

= -0.60, p = 0.552, OR = 0.91, 95% CI [0.66, 1.25]. A likelihood-ratio test evaluating the contribution of environment led to the same conclusion, $\chi^2(1) = 0.34$, p = 0.562. Disruptive disengagement was selected on 45.6% of quiet-condition trials and 44.5% of noisy-condition trials. H2 was not supported. Figure 2 displays the descriptive selection proportions in the two environments.

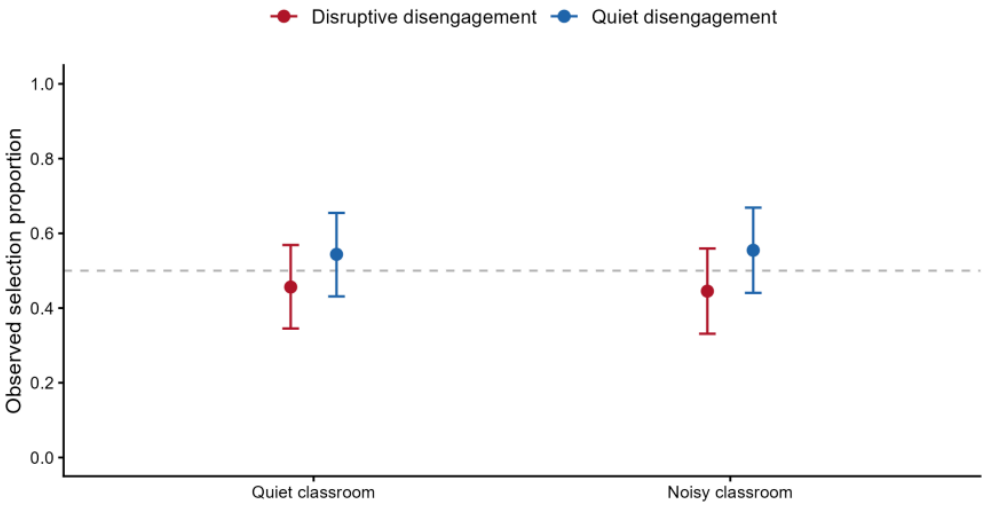

Figure 2. Final selection proportions by classroom environment

Note: Points are mean participant-level selection proportions (40 participants; 640 valid formal trials per environment). Error bars are pointwise 95% percentile bootstrap confidence intervals from 20,000 resamples of participants, retaining paired conditions within each resample (seed = 20261003). These are descriptive intervals, not model-adjusted probabilities or a test of the between-environment difference.

3.1.3. Secondary trial-level predictors

Three task-related predictors were associated with final selection. When the disruptive profile appeared on the right rather than the left, the odds of selecting it were lower, b = -0.37, SE = 0.17, z = -2.25, p = 0.025, OR = 0.69, 95% CI [0.50, 0.95]. Male-version stimuli were associated with lower odds of disruptive selection than female-version stimuli, b = -0.45, SE = 0.16, z = -2.74, p = 0.006, OR = 0.64, 95% CI [0.46, 0.88]. Trial order also predicted choice: later trials were associated with a greater likelihood of disruptive selection, b = 0.20, SE = 0.08, z = 2.43, p = 0.015, OR = 1.22, 95% CI [1.04, 1.43]. These effects were secondary to the primary hypotheses and were not adjusted as a family of confirmatory tests; they are therefore treated as task-related signals requiring replication. The fixed-effect estimates for this model are reported in Table 2.

Table 2. Binomial mixed-effects model predicting disruptive selection

Predictor	*b*	*SE*	*z*	*p*	*OR*	95% CI
Classroom environment: noisy	-0.10	0.16	-0.60	0.552	0.91	[0.66, 1.25]
Disruptive profile on right	-0.37	0.17	-2.25	0.025	0.69	[0.50, 0.95]
Stimulus-gender version: male	-0.45	0.16	-2.74	0.006	0.64	[0.46, 0.88]
Standardised trial order	0.20	0.08	2.43	0.015	1.22	[1.04, 1.43]

Note: Outcome coded 1 = disruptive selection and 0 = quiet selection. The Environment by Stimulus Gender interaction was tested first and was not significant, $\chi^2(1) = 0.19$, p = 0.665.

3.2. Temporal measures

The mean total response time was 3,351 ms in the quiet condition and 3,131 ms in the noisy condition. The adjusted noisy-minus-quiet contrast on log response time was -0.078, *SE* = 0.025, 95% CI [-0.127, -0.030], *p* = 0.002. This corresponds to a ratio of adjusted geometric means of 0.925. The Environment by Final Selection interaction was not significant, *p* = 0.877, nor was the equally weighted final-selection contrast, *p* = 0.289. Figure 3 displays the descriptive temporal measures, and Table 3 summarises the adjusted contrasts.

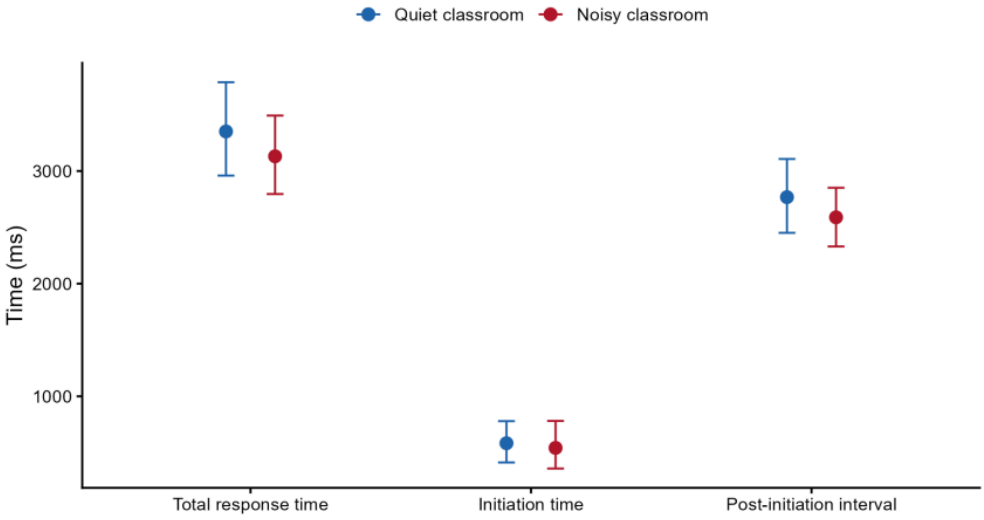

Figure 3. Temporal measures by classroom environment

Note: Points show mean participant-level times; error bars are pointwise 95% percentile intervals from 10,000 participant bootstrap resamples within each environment (seed = 20261007). They describe arithmetic means, whereas the models analyse log times. Separate mean intervals do not test the paired environment contrast.

Table 3. Temporal measures by classroom environment

Outcome	Quiet Mms	Noisy Mms	Environmentcontrast (SE)	95% CI	*p*	Interactionp
Total response time	3,351	3,131	-0.078 (0.025)	[-0.127, -0.030]	0.002	0.877
Initiation time	583	542	-0.062 (0.040)	[-0.141, 0.016]	0.120	0.536
Post-initiation interval	2,768	2,589	-0.075 (0.028)	[-0.131, -0.019]	0.008	0.494

Note: Environment contrasts are noisy minus quiet on the log scale, averaged equally over final-selection categories. They differ from coefficients at a single selection reference level. Means are untransformed milliseconds; interaction *p* values use Satterthwaite tests. Zero situation variance was removed from the initiation model.

Mean initiation time was 583 ms in the quiet condition and 542 ms in the noisy condition. The adjusted environment contrast was -0.062, *SE* = 0.040, 95% CI [-0.141, 0.016], *p* = 0.120. Neither the interaction, *p* = 0.536, nor the selection contrast, *p* = 0.649, was significant.

Mean post-initiation interval was 2,768 ms in the quiet condition and 2,589 ms in the noisy condition. Its adjusted environment contrast was -0.075, *SE* = 0.028, 95% CI [-0.131, -0.019], *p* = 0.008, corresponding to a

geometric-mean ratio of 0.927. The interaction, p = 0.494, and selection contrast, p = 0.150, were not significant. The interval includes motion, pauses, and final-click delay.

Allowing participant-specific environment slopes retained negative adjusted contrasts for total response time (p = 0.011) and post-initiation interval (p = 0.034). The unadjusted arithmetic mean differences were -221 ms, 95% participant bootstrap CI [-503, 45], and -179 ms, CI [-448, 70], respectively. These intervals concern a different scale and omit the model covariates; they limit any claim that the speed difference is insensitive to analysis choices. The random-slope sensitivity results and unadjusted participant-paired differences are reported in Appendix D (Tables D1 and D2, respectively).

3.3. Mouse-trajectory measures

Mean MAD was 0.650 (*SD* = 0.522); mean absolute net AUC was 0.348 (*SD* = 0.348). On the *log(1 + value)* scale, neither the adjusted environment contrast for MAD, -0.017, 95% CI [-0.046, 0.013], p = 0.279, nor that for AUC, -0.009, CI [-0.033, 0.014], p = 0.447, was significant. Selection contrasts were also nonsignificant (MAD p = 0.129; AUC p = 0.502), as were the interactions (MAD p = 0.597; AUC p = 0.699). The AUC situation variance was zero and its model used a participant intercept only. Table 4 reports the primary curvature-model contrasts.

Mean path efficiency was 0.622 (*SD* = 0.257). Its environment contrast (p = 0.085), selection contrast (p = 0.160), and interaction (p = 0.542) were nonsignificant. Efficiency was higher with the disruptive profile on the right (b = 0.029, p = 0.022) and on later trials (b = 0.019, p = 0.003). At 20% of the first-to-last recorded interval, horizontal position was toward the unselected option on 41.1% of quiet-choice trials and 38.3% of disruptive-choice trials. Maximum deviation occurred at a mean of 44.3% of that interval; 101-point paths averaged 2.53 horizontal direction changes. These implementation-dependent descriptive features were exploratory.

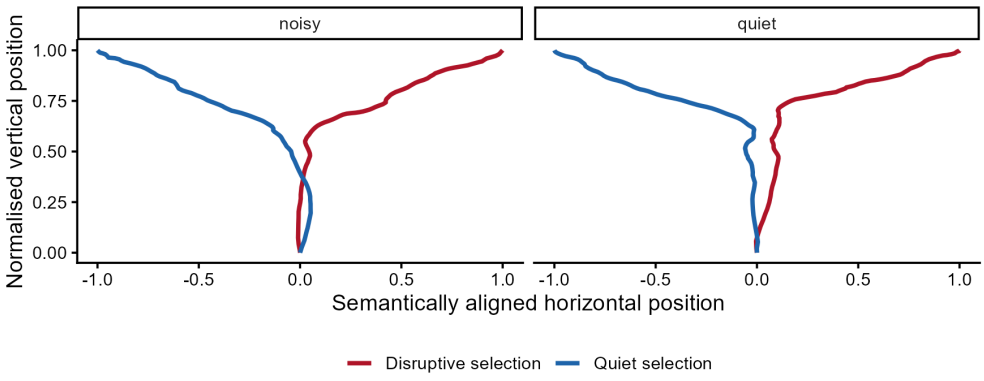

Figure 4. Semantically aligned mouse trajectories by classroom environment and final selection

Note: All 1,280 valid formal trials were used: quiet environment, 348 quiet and 292 disruptive choices; noisy environment, 355 quiet and 285 disruptive choices. Endpoints are normalised to (-1, 1) or (1, 1). Each line averages participant cell means after 101-point interpolation. Contributors were 38 and 32 participants for quiet and disruptive selections in the quiet environment, and 36 and 33 in the noisy environment; participants with no choice of a type do not contribute to that line.

Intervals are model-based 95% confidence intervals on the transformed scale. Full coefficients and model sensitivity checks are provided in Appendix D.

Table 4. Primary mouse-trajectory curvature models

Outcome	Contrast or term	Estimate (SE)	95% CI	*p*
MAD	Environment	-0.017 (0.015)	[-0.046, 0.013]	0.279
MAD	Final selection	-0.032 (0.021)	[-0.074, 0.009]	0.129
MAD	Environment × Selection	-0.016 (0.031)	—	0.597
AUC	Environment	-0.009 (0.012)	[-0.033, 0.014]	0.447
AUC	Final selection	-0.011 (0.016)	[-0.042, 0.021]	0.502
AUC	Environment × Selection	0.009 (0.024)	—	0.699

Note: MAD = Maximum Absolute Perpendicular Deviation; AUC = Absolute Net Geometric Polygon Area in Normalised Coordinates. Both outcomes use *log(1 + value)*. Environment and selection contrasts average equally over the other factor. AUC uses only a participant random intercept because situation variance was zero.

3.4. Exploratory drift-diffusion modelling

The common four-parameter model estimated v = -0.112, a = 3.297, relative z = 0.542, and t_0 = 0.615 s. The freely estimated starting point improved the pooled likelihood relative to z fixed at 0.50, LR $\chi^2(1) = 20.50$, nominal $p < 0.001$. Participant bootstrap intervals are shown in Table 6. The drift interval included zero, whereas the relative starting-point interval was above 0.50. These parameters summarise pooled response distributions; their combination does not establish an initial visual attraction to disruption followed by a within-person reversal.

Allowing drift, starting point, or non-decision time to vary by environment did not appreciably improve fit (nominal p = 0.507, 0.630, and 0.470, respectively). Allowing boundary separation to vary gave a small improvement, LR $\chi^2(1) = 3.90$, nominal p = 0.048, but q = 0.193 after adjustment across four parameter comparisons. AIC favoured this boundary model by only 1.90, whereas BIC favoured the common model by 3.26. No environment comparison survived the adjustment. These pooled likelihood comparisons do not provide participant-level causal or mechanistic tests. Table 5 compares the common, midpoint-fixed, and environment-specific models.

Table 5. Exploratory drift diffusion model comparison

Model	*k*	AIC	BIC	LR χ^2	Nominal *p*	*q*
Midpoint fixed $z = 0.50$	3	6,453.84	6,469.31	20.50	< 0.001	—
Common $v\ a\ z\ t_0$	4	6,435.34	6,455.96	—	—	—
Drift varies by environment	5	6,436.90	6,462.67	0.44	0.507	0.630
Boundary varies by environment	5	6,433.44	6,459.21	3.90	0.048	0.193
Starting point varies	5	6,437.11	6,462.88	0.23	0.630	0.630
Non-decision time varies	5	6,436.82	6,462.59	0.52	0.470	0.630

Note: Likelihoods were conditioned on RT < 12 s. Each model converged from all 12 initialisations to the same optimum within 0.001 negative-log-likelihood units. For the midpoint row, LR and p compare it with the common model; for environment rows

they compare each larger model with the common model. q values apply only to the four environment comparisons. Pooled nominal p values ignore participant clustering.

Table 6. Common diffusion parameters and participant bootstrap intervals

Parameter	Estimate	95% lower	95% upper
Drift rate v	-0.112	-0.257	0.042
Boundary separation a	3.297	3.031	3.577
Relative starting point z	0.542	0.507	0.572
Non-decision time t_0 s	0.615	0.539	0.704

Note: z is a relative fraction of boundary separation, with 0.50 indicating the midpoint. Intervals are percentile intervals from 200 successful participant bootstrap fits out of 200 resamples. Diffusion noise is fixed at $s = 1$. Parameter estimates are pooled and do not estimate individual teacher mechanisms.

The common model predicted a disruptive-choice probability of 0.451, compared with the observed 0.451. Across response-specific 0.10, 0.30, 0.50, 0.70, and 0.90 quantiles, the largest absolute RT discrepancy was 0.25 s. Appendix C reports these checks and the observed and fitted joint CDFs. The mismatch in parts of the RT distributions limits literal interpretation of the fitted accumulation parameters.

3.5. Exploratory participant-level differences

The final-selection model showed substantial between-participant variability. Participant-level random-intercept variance was 7.94 (SD = 2.82), compared with a situation-level variance of 0.23 (SD = 0.48). Eighteen participants selected disruptive disengagement on more than half of their trials and were classified as disruptive-preferring; 22 selected it on fewer than half and were classified as quiet-preferring. No participant selected each option exactly 16 times.

Table 7. Key exploratory participant-level comparisons

Variable	Disruptive-preferring	Quiet-preferring	Test	p / q
n	18	22	—	—
Disruptive selections, M %	83.0	14.1	Descriptive	—
Form-teacher experience, M years	3.72	0.77	U = 302	p = 0.003; q = 0.023
Any form-teacher experience	13/18 (72.2%)	8/22 (36.4%)	Fisher exact; OR ≈ 4.55	p = 0.031
Total teaching experience, M years	6.42	3.70	U = 268	p = 0.056

Note: Preference groups were defined post hoc from observed task choices. q is the Benjamini-Hochberg adjusted value for the form-teacher-years comparison.

Form-teacher experience differed between the groups. Disruptive-preferring participants reported more form-teacher experience (M = 3.72 years, *median* = 3) than quiet-preferring participants (M = 0.77 years, *median* = 0), U = 302, p = 0.003. The difference remained significant after Benjamini-Hochberg correction, q = 0.023. Having any previous form-teacher experience was also more common in the disruptive-preferring group (72.2%) than in the quiet-preferring group (36.4%), unadjusted Fisher's exact p = 0.031, cross-product

OR ≈ 4.55. Total teaching experience showed the same descriptive direction but did not reach significance, U = 268, p = 0.056. Other demographic variables were not significantly associated with preference-group membership. These post-hoc group comparisons are exploratory and should not be interpreted causally. Table 7 summarises the key participant-group comparisons.

4. Discussion

4.1. Overview of findings

Neither primary hypothesis was supported when both student profiles were explicitly available and the decision concerned support priority. The noisy environment did not significantly increase disruptive selections. Adjusted models showed shorter total response time and post-initiation intervals in the noisy environment, with the same direction in models allowing participant-specific environment slopes. The unadjusted arithmetic mean-difference intervals included zero. No significant environment effect was detected for initiation time, MAD, or absolute net AUC.

The exploratory pooled DDM estimated a starting point closer to the disruptive boundary and drift toward the quiet boundary. This describes two fitted parameter directions, rather than direct evidence that teachers first attended to disruption and then redirected attention. Participant-level uncertainty and imperfect RT-distribution fit restrict that interpretation, and no environment-specific parameter survived adjustment across the four comparisons.

4.2. What the null choice results constrain

The primary tests did not detect the disruptive-choice pattern predicted by H1 and H2 in this support-prioritisation task. The evidence is specific to decisions made after both profiles were displayed. It does not show that salience is irrelevant, that preferences are equivalent, or that small effects are absent.

Models of teacher competence and professional vision distinguish perception, interpretation, and decision making [2, 3, 20]. Eye-tracking studies examine visual sampling, whereas this task supplied both profiles and recorded a response priority. The results therefore concern explicit choice more directly than spontaneous detection or the content of teachers' interpretations. Quiet disengagement could still be overlooked in a live classroom even when it receives support priority once displayed.

The support-oriented wording offers one plausible account of the null disruptive advantage. A teacher deciding who is most disruptive may reasonably prioritise behaviour that interrupts the class, whereas a teacher deciding who needs support may assign greater value to withdrawal or low participation. The current study cannot establish that framing caused the observed pattern because no alternative prompt was tested. The results are instead consistent with the broader idea that task goals influence which available information becomes decision-relevant [11]. A direct test would require experimentally comparing support-first, disruption-first, and urgency-first instructions.

4.3. Classroom environment and response dynamics

The noisy classroom condition did not alter final prioritisation, but participants completed responses more quickly. The descriptive timing difference was concentrated in the post-initiation interval. This interval includes pauses and final-click delay as well as cursor motion, so the pattern does not isolate faster motor execution. Reviews of noise effects on cognition emphasise that moderate environmental noise can produce negative, null, or occasionally beneficial performance effects depending on task demands and individual

differences [13, 14]. Response strategy, arousal, response caution, and movement or clicking behaviour remain possible explanations that this design cannot separate. The timing association was clearer in adjusted log-scale models than in unadjusted arithmetic mean contrasts. It should therefore be described as a model-dependent association with response timing rather than an unqualified acceleration for every teacher.

The DDM provided only weak evidence for a lower boundary in the noisy environment: the nominal pooled test was $p = 0.048$, but $q = 0.193$ across four comparisons, and BIC favoured the common model. This does not establish reduced response caution. The model pooled heterogeneous teachers, and its RT quantile discrepancies further limit mechanistic interpretation. Because background and audio changed together and perceived demand was not measured, timing differences cannot be attributed uniquely to auditory noise or cognitive load.

4.4. Mouse trajectories and decision competition

Reconstructed MAD and absolute net AUC showed no reliable association with environment, final selection, or their interaction. Descriptive paths contained early movement toward the unselected option and direction changes, but these features do not establish stronger early attraction toward one disengagement type. Geometric measures were defined explicitly: absolute net AUC can conceal cancelling excursions, and endpoint-based spatial scaling and the recorded sampling interval affect the measures. Trajectory interpretation depends on task implementation and the movement processes generating the paths [21, 17].

Mouse trajectories describe recorded response geometry and timing, whereas the pooled diffusion model summarises choice and total RT through an accumulation model. The fitted combination of $z > 0.50$ and $v < 0$ distinguishes two model parameters. Its meaning is constrained by the participant bootstrap intervals and imperfect distributional fit; it does not establish the same process in individual teachers.

4.5. Task-related and individual differences

Secondary associations with profile position and trial order show that task structure also needs consideration. Disruptive profiles were selected less often when presented on the right, while later trials were associated with more disruptive selections. Position and trial order also predicted path efficiency. These associations support retaining layout and sequence covariates, while their psychological explanation remains unresolved [17, 22].

Stimulus-gender version predicted final choice, but this should not be treated as evidence of a stable student-gender bias. The male and female images were corresponding AI-generated sets rather than controlled morphs, so perceived gender was inseparable from possible differences in expression, posture, contrast, and other visual properties. More broadly, the disengagement stimuli themselves were not independently pretested for salience, severity, realism, or support need. The study therefore tests the operationalised quiet-versus-disruptive stimulus sets; it does not isolate perceptual salience as a single experimentally validated dimension.

Participant-level variability exceeded situation-level variability, and form-teacher experience was the clearest exploratory group difference. Teachers with greater responsibility for a class group might assign different urgency to disruption, but this interpretation was not tested directly. The preference groups were defined after observing the choices, were small (18 and 22 participants), and were not part of a causal or moderation design. The association should therefore be treated as a hypothesis for replication.

4.6. Theoretical contribution

The study separates a support-priority outcome from the broader concept of noticing. Its contribution is to show that the hypothesised disruptive-choice advantage was not detected when teachers compared two explicit

needs. This result identifies a boundary for that prediction within the implemented task, while leaving open whether cue detection, interpretation, and priority differ in natural classrooms.

4.7. Practical implications

Quiet disengagement was frequently selected when its cues were explicitly available, but the task does not show that such students receive adequate support in everyday classrooms. Training that helps teachers identify subtle disengagement could be evaluated alongside subsequent support decisions. Teacher-noticing interventions and classroom simulations provide settings for that work [23, 24]. The present study did not test a training intervention or its effects on student engagement.

4.8. Limitations and future directions

The sample of 40 teachers limits precision for small effects and participant-level differences. The paired-test planning benchmark does not establish power for the binary mixed-effects models. Participant random slopes for environment were assessed in sensitivity analyses, but several outcomes provided insufficient variance to estimate them reliably. The temporal findings also depended on transformation and covariate adjustment. Larger studies with model-specific power planning and preregistered analyses would strengthen inference.

Stimulus validity was limited. The two-teacher plausibility review did not provide independent quantitative ratings of salience, severity, or support need. Male and female images were separate generated sets rather than controlled morphs, so their contrast could reflect other visual differences. Future studies should pretest descriptions and images separately and distinguish salience from support need before testing their effects.

The classroom-environment contrast combined different visual scenes and audio files without a direct manipulation check. Exposure began before the start-button click, making pre-choice exposure duration participant-dependent. Playback success and sound pressure level were not recorded. Faster responses cannot therefore be assigned to sound, visual complexity, or perceived workload. Independent manipulation of these components and measurement of perceived demand would clarify the effect.

Timed-out attempts were excluded and repeated until a valid response was obtained. This produced a complete set of final choices but exposed a subset of participants to repeated trials. Future studies should specify how first attempts, timeouts, and repeated exposures enter the analysis. Cursor recording began with the first mousemove event and did not add the start or final click coordinates; the recorded path therefore need not encompass the entire physical response. Endpoint normalisation cannot recover unrecorded positions, and its unequal horizontal and vertical scaling changes geometric distances.

Secondary temporal and trajectory analyses involved multiple tests without a family-wide adjustment, and residual plots showed departures from normality for some outcomes, especially initiation time and AUC. Their p values should be interpreted cautiously. Final selection was observed rather than randomised; conditional process-model associations are not causal effects of choice. DDM fits were exploratory and pooled, with no teacher or situation parameters. Bootstrap intervals represented participant variation for common parameters, but the pooled likelihood-ratio tests ignored that dependence. Conditioning on the deadline addressed truncation of valid responses but did not model repeated exposure or the behavioural history of timeouts. RT-distribution mismatch further limits the model as an account of individual cognition.

Finally, the task was a structured situational judgement rather than direct observation of classroom noticing [25]. Both profiles were displayed and the prompt specified support priority. Studies that vary the prompt and combine choice tasks with classroom video or eye tracking could separate detection from response selection [4, 5, 6]. Such designs would also help test whether the proposed support-goal interpretation accounts for the current pattern.

5. Conclusion

In an explicit support-prioritisation task, neither an overall disruptive-choice advantage nor an increase in disruptive choices in the noisy environment was detected. Adjusted analyses associated the noisy condition with shorter total response time and post-initiation intervals, while unadjusted mean-difference intervals included zero. Reconstructed trajectory measures showed no reliable environment or choice effects. The exploratory pooled DDM described a starting-point and drift asymmetry but provided no adjusted evidence of environment-specific parameter differences, with uncertainty and fit limitations restricting mechanistic interpretation. These results concern choices between two displayed profiles and do not establish equivalent preferences or spontaneous noticing in live classrooms.

References

[1] Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–20. https://doi.org/10.1146/annurev.ps.41.020190.000245

[2] Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

[3] König, J., Santagata, R., Scheiner, T., Adleff, A.-K., Yang, X., & Kaiser, G. (2022). Teacher noticing: A systematic literature review of conceptualizations, research designs, and findings on learning to notice. *Educational Research Review, 36*, Article 100453. https://doi.org/10.1016/j.edurev.2022.100453

[4] Weyers, J., König, J., Santagata, R., Scheiner, T., & Kaiser, G. (2023). Measuring teacher noticing: A scoping review of standardized instruments. *Teaching and Teacher Education, 122*, Article 103970. https://doi.org/10.1016/j.tate.2022.103970

[5] Witt, J., Schorer, J., Loffing, F., & Roden, I. (2024). Eye-tracking research on teachers' professional vision: A scoping review. *Teaching and Teacher Education, 144*, Article 104568. https://doi.org/10.1016/j.tate.2024.104568

[6] Keskin, Ö., Seidel, T., Stürmer, K., & Gegenfurtner, A. (2024). Eye-tracking research on teacher professional vision: A meta-analytic review. *Educational Research Review, 42*, Article 100586. https://doi.org/10.1016/j.edurev.2023.100586

[7] Stahnke, R., & Blömeke, S. (2021). Novice and expert teachers' situation-specific skills regarding classroom management: What do they perceive, interpret and suggest? *Teaching and Teacher Education, 98*, Article 103243. https://doi.org/10.1016/j.tate.2020.103243

[8] Vargas-Madriz, L. F., Konishi, C., & Wong, T. K. (2024). A meta-analysis of the association between teacher support and school engagement. *Social Development, 33*(4), Article e12745. https://doi.org/10.1111/sode.12745

[9] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

[10] Lawson, H. A., & Lawson, M. A. (2020). Student engagement and disengagement as a collective action problem. *Education Sciences, 10*(8), Article 212. https://doi.org/10.3390/educsci10080212

[11] Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology, 29*, 97–101. https://doi.org/10.1016/j.copsyc.2018.12.024

[12] Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science, 44*(3), 243–265. https://doi.org/10.1007/s11251-016-9367-z

[13] Sörqvist, P. (2014). On interpretation and task selection in studies on the effects of noise on cognitive performance. *Frontiers in Psychology, 5*, Article 1249. https://doi.org/10.3389/fpsyg.2014.01249

[14] Zhou, H., Molesworth, B. R. C., Burgess, M., & Hatfield, J. (2024). The effect of moderate broadband noise on cognitive performance: A systematic review. *Cognition, Technology & Work, 26*(1), 1–36. https://doi.org/10.1007/s10111-023-00746-2

[15] Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

[16] Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (2nd ed., pp. 111–130). Routledge. https://doi.org/10.4324/9781315160559-9

[17] Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2025). Movement tracking of psychological processes: A tutorial using mousetrap. *Behavior Research Methods, 57*, Article 307. https://doi.org/10.3758/s13428-025-02695-2

[18] Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

[19] Singmann, H., Brown, S., Gretton, M., Heathcote, A., & Fernandez, K. (2026). rtdists: Response time distributions (Version 0.12-0) [R package]. https://doi.org/10.32614/CRAN.package.rtdists

[20] van Es, E. A., & Sherin, M. G. (2021). Expanding on prior conceptualizations of teacher noticing. *ZDM-Mathematics Education, 53*(1), 17–27. https://doi.org/10.1007/s11858-020-01211-4

[21] Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences, 22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

[22] Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods, 50*(5), 2097–2110. https://doi.org/10.3758/s13428-017-0977-4

[23] Stürmer, K., Fütterer, T., Kron, S., Sommerhoff, D., & Ufer, S. (2024). What makes a simulation-based learning environment for preservice teachers authentic? The role of individual learning characteristics and context-related features. *European Journal of Psychology of Education, 39*(4), 3277–3299. https://doi.org/10.1007/s10212-024-00837-2

[24] Telgmann, L., & Müller, K. (2023). Training & prompting pre-service teachers' noticing in a standardized classroom simulation: A mobile eye-tracking study. *Frontiers in Education, 8*, Article 1266800. https://doi.org/10.3389/feduc.2023.1266800

[25] Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology, 9*(1), 3–22. https://doi.org/10.1017/iop.2015.71

Appendix

Appendix A. Supplementary diagnostic plot

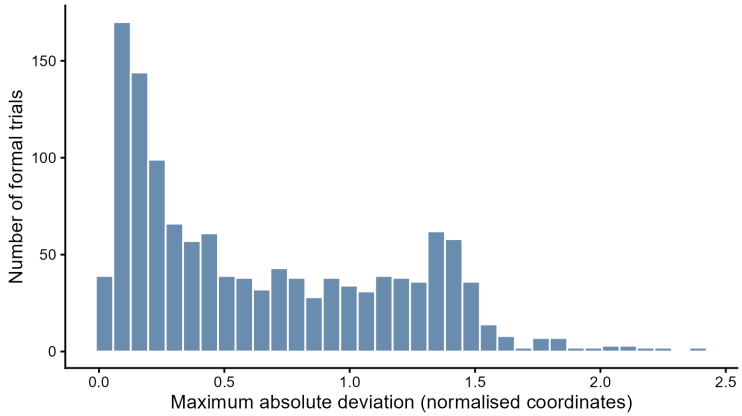

Figure A1. Distribution of reconstructed maximum absolute deviation

Note: All 1,280 valid formal trials are included. MAD is calculated from spatially normalised recorded paths; no practice or timed-out attempts are included. The histogram is descriptive, and the model uses *log(1 + MAD)*.

Appendix B. AI generated stimuli and retained generation records

The implemented set comprised 32 classroom images, organised as four images per situation: male disruptive, male quiet, female disruptive, and female quiet. Figures B1–B4 show examples. The retained Gemini conversation shows an iterative process: an initial portrait-style proposal was rejected in favour of classroom scenes; subsequent requests specified corresponding views and behaviour-by-gender combinations. The classroom images show differences in expression, posture, gaze, framing, and background. Requests for consistent views do not establish that these features were experimentally matched.

Figure B1. Male-version disruptive-disengagement example

Figure B2. Male-version quiet-disengagement example

Figure B3. Female-version disruptive-disengagement example

Figure B4. Female-version quiet-disengagement example

The excerpts below provide English translations of retained user prompts from the generation conversation. They document part of the generation process, not a complete reproducible prompt history. The exact model version, generation dates, seeds, and complete sequence of inputs are unavailable. The implemented English behavioural descriptions are reproduced separately so that the actual task materials can be distinguished from earlier conversational suggestions.

(1) Retained user prompt excerpts in English translation

1) The images should show different states of a student viewed from the front in a classroom setting.

2) Use the same view for group 1: male disruptive at the upper left, male quiet at the upper right, female disruptive at the lower left, and female quiet at the lower right.

3) Create four images based on the supplied image.

4) Create four images for the second group based on the descriptions above.

5) Create two four-panel student images for groups 3 and 4.

(2) Implemented behavioural descriptions

The following strings were extracted from the formal task code. Each pair was reused for male and female versions and both classroom environments.

Situation A

Disruptive: Giggle with a peer and grab their stationery, ignoring the ongoing lecture.

Quiet: Stare blankly out the window while spinning a pen, lost in a private daydream.

Situation B

Disruptive: Close the textbook loudly and push the desk away in a clear show of frustration.

Quiet: Rest the head on arms with the book on an old page, showing no sign of following along.

Situation C

Disruptive: Stand up abruptly and noisily move the chair, disrupting the classroom silence.

Quiet: Scribble on the book margin without eye contact, avoiding any interaction with the teacher.

Situation D

Disruptive: Tap a ruler on the desk to draw peers' attention, waiting for someone to react.

Quiet: Gaze vacantly at the blackboard without taking notes, remaining unresponsive to the lesson.

Situation E

Disruptive: Share snacks loudly with a peer during the lecture, completely distracted from the task.

Quiet: Fixate on a book corner with a confused look, failing to respond when the teacher calls.

Situation F

Disruptive: Intentionally drop a pen to make noise and bend over, looking around to see who is watching.

Quiet: Huddle in a large hoodie with the face half-hidden, trying to disappear into the seat.

Situation G

Disruptive: Carve on the desk while mimicking the teacher's voice, amusing the nearby students.

Quiet: Rearrange stationery repeatedly without looking up, absorbed in a meaningless repetitive task.

Situation H

Disruptive: Flip pages loudly and complain about the workload, making sure the teacher hears the sigh.

Quiet: Lean against the windowsill with a still pen on paper, completely disengaged from the writing task.

Appendix C. Diffusion fit checks

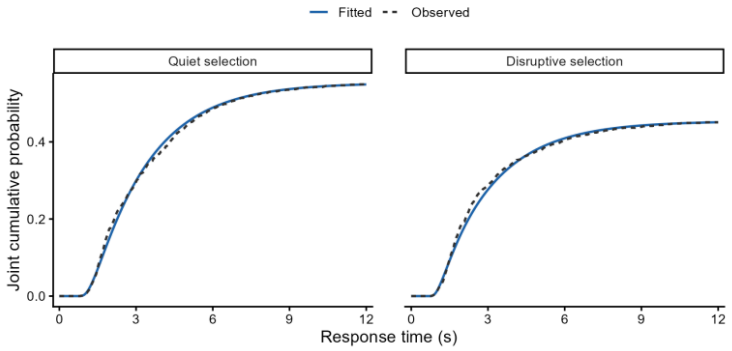

Figure C1. Observed and fitted joint response time distributions

Table C1. Choice conditional response time quantiles in seconds

Selection	Quantile	Observed	Fitted	Fitted—observed
Quiet	0.10	1.419	1.409	-0.011
Quiet	0.30	1.899	2.044	0.144
Quiet	0.50	2.796	2.796	0.000
Quiet	0.70	4.145	3.897	-0.248
Quiet	0.90	6.222	6.186	-0.037
Disruptive	0.10	1.278	1.224	-0.054
Disruptive	0.30	1.682	1.768	0.086
Disruptive	0.50	2.274	2.464	0.190
Disruptive	0.70	3.372	3.539	0.168
Disruptive	0.90	6.001	5.835	-0.167

The observed and fitted joint cumulative probabilities combine choice frequency and response time. The fitted common model conditions predictions on RT below 12 seconds. Curve agreement is descriptive and

does not validate an individual cognitive mechanism. Figure C1 plots the observed and fitted joint response-time distributions, and Table C1 reports the corresponding choice-conditional quantiles.

The fitted disruptive probability is 0.451. The observed proportion is .451. Quantile discrepancies indicate imperfect fit to the shape of the response distributions. Table C2 reports the environment-specific estimates from the separate one-parameter models.

Table C2. Environment specific estimates from separate one parameter models

Varying parameter	Quiet	Noisy	Nominal *p*	*q*
v	-0.101	-0.124	0.507	0.630
a	3.369	3.224	0.048	0.193
z	0.546	0.539	0.630	0.630
t0	0.626	0.603	0.470	0.630

For comparison, the untruncated common-model fit estimated v = -0.115, a = 3.250, z = 0.543, and t_0 = 0.626 s. All environment models in Table C2 vary only the named parameter. The remaining parameters are shared, so rows must not be combined into a model in which all four parameters vary simultaneously.

Appendix D. Model specification and sensitivity

Reference levels are quiet environment, quiet final selection, disruptive profile on the left, and female stimulus version. Trial order is standardised across retained trials. The coefficient tables below use these reference levels; the main-text contrasts average equally over the other interaction factor. Participant and situation intercepts are included except that the situation term is removed for initiation time and AUC when its variance is zero. Full fixed-effect coefficients are reported for total response time (Table D3), initiation time (Table D4), post-initiation interval (Table D5), MAD (Table D6), AUC (Table D7), and path efficiency (Table D8).

Table D1. Environment contrasts with participant random slopes

Outcome	Contrast	95% CI	*p*	Fit
Total response time	-0.078	[-0.137, -0.019]	0.011	Converged
Initiation time	-0.062	[-0.141, 0.016]	0.120	Zero slope variance
Post-initiation interval	-0.074	[-0.143, -0.006]	0.034	Converged
MAD	-0.017	[-0.046, 0.013]	0.279	Zero slope variance
AUC	-0.009	[-0.033, 0.014]	0.447	Zero slope variance

Table D2. Unadjusted participant paired arithmetic mean differences

Outcome	Noisy—quiet	95% lower	95% upper
Total response time	-220.7	-503.3	45.1
Initiation time	-41.6	-133.3	53.8
Post-initiation interval	-179.1	-448.0	69.8
MAD	-0.016	-0.062	0.030
AUC	-0.011	-0.041	0.021

Only total response time and post-initiation interval supported nonzero participant slope variance. The other slope models collapsed to the intercept model and are not independent confirmatory analyses. Table D1 reports the environment contrasts and convergence results for these sensitivity models.

Times are milliseconds; trajectory measures use normalised coordinates. Intervals use 20,000 participant bootstrap resamples of paired differences (seed = 20261008). They omit model covariates and concern untransformed arithmetic means. Table D2 reports these unadjusted participant-paired differences and bootstrap intervals.

Table D3. Fixed effects for Total response time

Term	b	SE	df	*t*	*p*
Intercept	7.962	0.066	67.1	120.948	< 0.001
Noisy environment	-0.082	0.034	1,228.3	-2.455	0.014
Disruptive selection	-0.042	0.044	1,270.0	-0.961	0.337
Disruptive profile on right	-0.007	0.025	1,234.5	-0.294	0.769
Male stimulus version	0.013	0.025	1,227.7	0.521	0.603
Standardised trial order	-0.170	0.012	1,232.4	-13.589	< 0.001
Environment × Selection	0.008	0.050	1,229.8	0.154	0.877

Table D4. Fixed effects for Initiation time

Term	*b*	*SE*	*df*	*t*	*p*
Intercept	5.462	0.171	46.3	31.914	< 0.001
Noisy environment	-0.037	0.054	1,234.2	-0.696	0.487
Disruptive selection	0.051	0.071	1,257.6	0.726	0.468
Disruptive profile on right	0.042	0.040	1,235.2	1.028	0.304
Male stimulus version	0.071	0.040	1,234.3	1.777	0.076
Standardised trial order	-0.155	0.020	1,234.2	-7.776	< 0.001
Environment × Selection	-0.050	0.080	1,234.3	-0.618	0.536

Table D5. Fixed effects for Post-initiation interval

Term	*b*	*SE*	*df*	*t*	*p*
Intercept	7.780	0.065	75.2	119.053	< 0.001
Noisy environment	-0.095	0.038	1,228.4	-2.478	0.013
Disruptive selection	-0.078	0.050	1,245.0	-1.575	0.115
Disruptive profile on right	-0.019	0.029	1,236.3	-0.671	0.502
Male stimulus version	-0.007	0.028	1,228.0	-0.244	0.807
Standardised trial order	-0.157	0.014	1,232.6	-11.002	< 0.001
Environment × Selection	0.039	0.057	1,230.2	0.683	0.494

Table D6. Fixed effects for MAD

Term	*b*	*SE*	*df*	*t*	*p*
Intercept	0.495	0.031	89.3	16.150	< 0.001
Noisy environment	-0.008	0.020	1,229.4	-0.411	0.681
Disruptive selection	-0.024	0.026	1,191.7	-0.929	0.353
Disruptive profile on right	-0.045	0.015	1,240.1	-2.921	0.004
Male stimulus version	-0.002	0.015	1,228.4	-0.153	0.878
Standardised trial order	-0.011	0.008	1,234.8	-1.436	0.151
Environment × Selection	-0.016	0.031	1,232.0	-0.529	0.597

Table D7. Fixed effects for AUC

Term	*b*	*SE*	*df*	*t*	*p*
Intercept	0.292	0.020	149.0	14.833	< 0.001
Noisy environment	-0.014	0.016	1,235.7	-0.855	0.393
Disruptive selection	-0.015	0.020	1,005.5	-0.771	0.441
Disruptive profile on right	-0.019	0.012	1,247.3	-1.567	0.117
Male stimulus version	-0.001	0.012	1,236.2	-0.101	0.920
Standardised trial order	-0.001	0.006	1,235.8	-0.102	0.919
Environment × Selection	0.009	0.024	1,237.7	0.387	0.699

Table D8. Fixed effects for path efficiency

Term	*b*	*SE*	*df*	*t*	*p*
Intercept	0.589	0.027	83.4	22.154	< 0.001
Noisy environment	0.014	0.017	1,229.0	0.829	0.407
Disruptive selection	0.017	0.022	1,219.5	0.796	0.426
Disruptive profile on right	0.029	0.013	1,238.6	2.294	0.022
Male stimulus version	0.000	0.013	1,228.2	0.020	0.984
Standardised trial order	0.019	0.006	1,234.2	3.007	0.003
Environment × Selection	0.015	0.025	1,231.4	0.610	0.542

Residual Q–Q plots showed visible departures from normality, particularly for initiation time and AUC, so their model tests remain approximate. Model coefficients and sensitivity checks are reported to make those limitations assessable.