INPLASY

Measuring cognitive load during learning using eye-tracking data: A systematic review protocol

INPLASY202690088

doi: 10.37766/inplasy2026.9.0088

Received: 21 September 2026

Published: 21 September 2026

Yang, JW; Lee, J; Kim, H.

Corresponding author:
Ji Won Yang

yangji921@snu.ac.kr

Author Affiliation:
Seoul National University.

ADMINISTRATIVE INFORMATION

Support - No support.

Review Stage at time of this submission - Data extraction.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202690088

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 21 September 2026 and was last updated on 21 September 2026.

INTRODUCTION

Review question / Objective How are eye-tracking measures used to assess cognitive load during learning, and what evidence supports their ability to distinguish intrinsic, extraneous, and germane cognitive load?

The review examines: (1) how cognitive load is defined, theoretically grounded, and experimentally operationalized; (2) which gaze and pupil measures are used, and how they are recorded and processed; (3) how these measures relate to self-reported load, learning or task performance, prior knowledge, and other physiological measures; and (4) whether evidence supports sensitivity to overall task demands, differentiation of cognitive load types, or both.

The population comprises learners undertaking educational or training tasks. The exposure of interest is experimental variation in task demands or instructional conditions intended to affect cognitive load. Comparisons are between manipulated conditions within or between participants. Outcomes are eye-tracking data and their relationships with other indicators of cognitive load and learning. The aim is to identify the interpretive strengths and limitations of eye-tracking methods in measuring cognitive load and inform their use in educational research.

Rationale Cognitive load theory links learning to the limited capacity of working memory. Distinguishing demands associated with task complexity, instructional presentation, and learning-related processing is relevant to instructional design, but these demands can be difficult to separate empirically. Self-report measures provide learners' perceptions of effort or difficulty, whereas performance measures describe task or learning outcomes. Neither necessarily identifies the timing or type of cognitive load experienced during learning.

Eye tracking provides temporally detailed information about gaze allocation and pupil responses. However, changes in these measures may also reflect visual properties, attention, reading processes, prior knowledge, or task strategy. The same direction of change may therefore have different interpretations across tasks and instructional contexts.

This review focuses on the chain linking experimental manipulation, the cognitive load construct, eye-tracking measurement, and supporting validation evidence. It will distinguish measures that respond to task demands from evidence that differentiates specific load types, and examine how recording and preprocessing choices affect interpretation. It complements the broader physiological-workload review by Das Chakladar and Roy (2024) and educational eye-tracking review by Liu and Cui (2025) by focusing on cognitive load operationalization and type-specific measurement evidence.

Condition being studied Cognitive load during learning and educational task performance. This is an educational and psychological construct, rather than a disease or clinical condition. The review considers overall cognitive load and the intrinsic, extraneous, and germane load distinctions used in the included literature. Intrinsic load concerns task element interactivity in relation to learner knowledge; extraneous load concerns demands associated with instructional presentation; germane load concerns processing directed toward learning and schema construction. Authors' definitions and theoretical frameworks will be retained because these distinctions are not used uniformly across studies.

METHODS

Search strategy An initial search was conducted in Web of Science, PsycINFO, MEDLINE, and ERIC for records available up to July 2026 (starting from 2016 - 10 years). The search combined three concept groups: eye tracking, educational context, and cognitive load. The conceptual search string was:

("eye-tracking" OR "eye tracking" OR "eye gaze" OR "eye movement*") AND (learning OR education) AND ("cognitive load" OR "mental effort" OR "working memory")

This expression summarizes the concepts reported in the manuscript; exact database-specific syntax, field tags, controlled vocabulary, platforms, search dates, and limits will be supplied from the original search histories. English-language, peer-reviewed journal articles are eligible.

Search results were imported into Rayyan for record management, deduplication, and screening. Any updated searches will be documented separately with their dates and revised search histories.

Participant or population Human participants undertaking learning, educational assessment, or training tasks, including school pupils, undergraduate and graduate students, and professional learners or experts in educational comparisons. Eligible studies concern nonclinical populations and participants able to complete the visual task. Clinical populations recruited specifically because of physical or psychological disorders are excluded. No fixed age limit is proposed. Vision eligibility and screening procedures will be extracted as reported, including normal or corrected-to-normal vision where specified.

Intervention Experimental manipulation of learning-related task demands or instructional conditions intended to alter cognitive load. Manipulations may vary task difficulty, element complexity or interactivity, or categorical instructional features such as modality, layout, signaling, redundancy, or content relevance. Eye tracking is the measurement method. Measured individual differences, such as expertise or prior knowledge, may be examined alongside the manipulation but do not alone constitute an eligible intervention.

Comparator An alternative experimentally defined task-demand or instructional condition, compared within participants, between groups, or in a mixed design. Examples include lower versus higher task demands and contrasting instructional formats. A no-treatment comparator is not required. Comparisons based only on naturally occurring expertise, performance, or post hoc gaze groups are insufficient in the absence of an eligible manipulation.

Study designs to be included Quantitative experimental and quasi-experimental studies using within-subjects, between-subjects, or mixed designs, with an experimentally defined manipulation relevant to cognitive load and eye-tracking measurement during an educational or training task. Purely observational or correlational studies without such a manipulation are excluded.

Eligibility criteria Include full-text, English-language, peer-reviewed journal articles that: (1) fall within the confirmed publication-date window; (2) involve human learners in an educational, assessment, or training context; (3) experimentally vary a task or instructional feature relevant to cognitive load; and (4) report gaze or pupil measures used to investigate cognitive load or learning-related mental effort.

Studies need not explicitly name an intrinsic, extraneous, or germane load type. The target construct must be identifiable from the research question, theoretical rationale, or interpretation, rather than inferred solely from the use of an eye tracker. Theoretical framing and explicit specification of load types will be coded separately.

Exclude reviews, book chapters, conference-only reports, non-peer-reviewed material, nonhuman studies, clinical populations as defined above, noneducational tasks, studies without a relevant manipulation, and reports that mention eye tracking without reporting relevant measurements. Exclude algorithm-development studies when their sole purpose is prediction or classification without an interpretable educational cognitive-load investigation.

Screen-based, head-mounted, and webcam-based systems are eligible if other criteria are met. Wearable devices, VR, AR, or game-based delivery alone is considered mostly out of scope for this review. Multiple reports from overlapping samples will be linked to avoid double counting.

Information sources Web of Science, PsycINFO, MEDLINE, and ERIC. The initial search covered records available up to March 2025. Database platforms, exact dates, and any supplementary sources or updated searches will be documented in the final search record. Eligible reports are restricted to peer-reviewed journal articles.

Main outcome(s) The primary outcomes are: (1) the direction and magnitude, where reported, of gaze and pupil responses to manipulated learning-related demands; and (2) evidence that these responses distinguish intrinsic, extraneous, or germane load, rather than only overall cognitive load.

Measures include fixation counts and durations, dwell time and area-of-interest allocation, saccades and transitions, pupil diameter, baseline-corrected pupil dilation, and other reported gaze or pupil indices. For each relevant result, the measure definition, units, task condition, recording phase, analysis window, and baseline will be recorded where available.

Reported between-condition differences, correlations, regression coefficients, standardized effects, confidence intervals, and statistical uncertainty will be extracted when available. Evidence will be interpreted alongside the manipulation, authors' theoretical claims, and accompanying load or performance measures. A change in an eye-tracking indicator alone will not be treated as proof that a specific cognitive load type was isolated.

Additional outcome(s) Relationships between eye-tracking measures and self-reported mental effort or cognitive load; learning or task performance, including accuracy, response time, retention, or transfer; prior knowledge or expertise; and other physiological measures. Additional methodological outcomes include calibration, sampling rate, recording duration and phase, visual and lighting conditions, data-quality handling, pupil baseline correction, and reporting completeness. Measurements will be described at the time points used in each study; self-report or test outcomes collected after a task will be distinguished from concurrent eye-tracking recordings.

Data management Search records are managed in Rayyan and extraction data in a structured Excel workbook. The extraction framework records study and sample characteristics, cognitive load definitions and theoretical framing, experimental manipulations, study design, eye-tracking apparatus and calibration, recording conditions and timing, gaze and pupil metrics, preprocessing, accompanying measures, statistical methods, and findings.

For the remaining verification, two reviewers will independently confirm full-text eligibility, verify extracted data against source articles, and resolve disagreements by discussion. Decisions and reasons for full-text exclusion will be retained. The actual reviewer workflow used for the earlier title/abstract screening and extraction stages will be reported separately.

Raw source descriptions will be retained alongside coded categories. Missing reporting will be coded NR; inapplicable fields will be coded NA. Ambiguous classifications will be flagged for adjudication. Authors' explicit load-type claims will be kept separate from reviewer-inferred classifications, and overall or unspecified load will remain a distinct category.

AI-assisted preliminary extraction, if used, will be documented by tool, version where available, task, and human verification procedure. No AIgenerated extraction will be treated as verified without checking the primary article.

Quality assessment / Risk of bias analysis A preliminary methodological quality assessment used the 14-item quantitative QualSyst checklist (Kmet, Lee, and Cook, 2004). Two reviewers assessed two reference studies identified as representing the lowest and highest journal impact factors in the set considered.

Items addressing design, participant selection, measurement, analysis, and reporting were rated 2 (met), 1 (partially met), or 0 (not met). The recorded assessments used all 14 items and a denominator of 28. Yulianandra et al. scored 20/28 (71.43%); Korbach et al. scored 23/28 (82.14%). Applicability judgments require confirmation before finalizing these preliminary scores.

These assessments and journal impact factor information informed informal judgments about other studies' acceptability. The observed score range does not establish a validated inclusion threshold or the quality of unassessed studies. Impact factor is a journal-level characteristic and cannot establish individual study quality. The influence of these judgments on inclusion will be documented.

Any subsequent appraisal will be documented separately. These preliminary assessments do not establish the risk of bias of all included studies.

Strategy of data synthesis A structured narrative synthesis with descriptive tables will examine the relationships among cognitive load operationalization, eye-tracking measurement, and validation evidence. Meta-analysis is not currently planned because the review question concerns measurement interpretation across heterogeneous constructs, manipulations, metrics, and tasks.

Study characteristics will be summarized using frequencies and appropriate descriptive statistics. Tables will cross-classify overall or type-specific load, explicit versus inferred specification, manipulation type, study design, gaze versus pupil measurement, and accompanying measures. All percentages will state their denominators, and overlapping categories will be labelled.

Findings will be organized by eye-tracking metric and cognitive load construct. Direction, magnitude, precision, task phase, and measurement conditions will be considered together. Statistically nonsignificant findings will be described without equating them with proof of no effect. Counts of significant findings will not be used as a substitute for assessing effect estimates, study design, and measurement validity.

Claims about differentiating load types will give particular attention to explicit theoretical definitions, separable or factorial manipulations, type-specific accompanying measures, and controls for plausible alternative explanations. Studies assessing overall load will not be relabelled as type-specific solely on the basis of reviewer inference.

The review will be reported using PRISMA 2020. Funnel plots and statistical tests of publication bias are not planned without a suitable quantitative synthesis. Publication and selective-reporting limitations will be discussed, including the restriction to English-language journal articles. Any later decision to conduct meta-analysis will require a documented amendment specifying the questions, metrics, dependence handling, and model.

Subgroup analysis Descriptive stratified comparisons will be made, where sufficient studies are available, by: cognitive load type or combination versus overall/unspecified load; explicit versus reviewer-inferred specification; gaze, pupil, or combined measurement; difficulty, complexity, categorical, or combined manipulation; within-subjects, between-subjects, or mixed design; educational domain and learner population; and apparatus or recording context.

These are exploratory narrative and tabular comparisons, not statistical meta-analytic subgroup tests. Sparse categories will be described without claims of comparative superiority. Apparent differences will be interpreted in light of task, sample, and measurement differences.

Sensitivity analysis Robustness will be examined narratively by comparing the main conclusions with restricted syntheses that: (1) retain only studies with explicit, theoretically grounded load-type specification; (2) remove studies with unresolved eligibility or manipulation concerns; and (3) set aside findings whose interpretation depends on major identified design or measurement limitations, such as uncontrolled luminance in pupil comparisons or inadequately described data-quality handling.

Borderline cases and the reasons for restrictions will be recorded before the final reanalysis, together with the date and whether relevant findings were already known. Differences in the available evidence and conclusions will be reported. Statistical leave-one-out analysis is not planned without meta-analysis.

Language restriction Only English-language publications are eligible. The review and registration will be written in English.

Country(ies) involved Republic of Korea.

Other relevant information This review is being registered retrospectively. Database searching, study screening has been undertaken and data extraction is currently ongoing.
The record describes the methods used to date and the plans for remaining work. Study findings were available when this registration draft was prepared. Decisions made or revised after findings became available, including eligibility clarification, coding, appraisal, or synthesis decisions, will be identified with their timing and rationale. This registration does not establish prospective prespecification of earlier decisions.

Keywords cognitive load; eye tracking; learning; systematic review.

Dissemination plans The findings will be prepared for submission to a peer-reviewed journal. The final report will include the INPLASY registration identifier, registration timing, and any substantive methodological amendments.

Contributions of each author

Author 1 - Ji Won Yang - Conceptualization, methodology, literature searching, study selection, data extraction, synthesis, and writing the original draft.
Email: yangji921@snu.ac.kr
Author 2 - Jihyun Lee - Conceptualization, methodological supervision, interpretation of findings, and critical review and editing of the manuscript.
Email: leeji1@snu.ac.kr
Author 3 - Heeyoung Kim - Data extraction, synthesis, and review of the manuscript.
Email: young7481@snu.ac.kr