JSER

Journal of Science Education Research
Journal homepage: https://journal.uny.ac.id/index.php/jser/index

ISSN: 2597-9701

Visual Design in High-Stakes Science Assessments: A Longitudinal Content Analysis of Singapore Primary Science Examinations

Wai Mun Loke[1*]
[1] *Innovprof Singapore, 27 Orange Grove Road, #05-02, Singapore 258356.*
[*] *Corresponding Author. Email: wai.mun.loke@innovprof.com Phone*: +65 9437 8166*

Abstract
This study investigates the structural and functional design of visual representations within high-stakes primary science assessments. A total-population content analysis was conducted on Singapore Primary School Leaving Examination (PSLE) Science papers across a three-year window from 2023 to 2025. The evaluated corpus examined every visual representation deployed across all 120 test items ($N_{MCQ} = 84$; $N_{SAQ} = 36$). Guided by visual semiotic and cognitive frameworks, 134 extracted visual units were categorized into four structural typologies (Representational, Abstract, Data-Driven, Relational) and three functional levels (Essential, Supporting, Decorative). Observational coding revealed an absolute absence of decorative visuals (0.00%), with 97.01% (n = 130) categorized as functionally essential for item completion. A Pearson's Chi-Square Test of Independence demonstrated that the deployment of visual typologies was significantly dependent upon the specific scientific theme ($\chi^2(12) = 26.76$, $p = 0.0084$), with the Cycles thematic block containing 100% of all relational graphics. Longitudinal tracking indicated a distributional shift in visual deployment across question formats: visual elements in multiple-choice items decreased from 39 in 2023 down to 19 in 2025, whereas representational graphics within short-answer items increased from 6 to 11 over the same period. These results document empirical patterns of visual integration across item formats and domain topics, providing descriptive baselines for visual design in standardized STEM testing environments.

History
Received: June 24, 2026
Revised: August 29, 2026
Accepted: September 1, 2026

Keywords:
Assessment design, Cognitive load theory, Multimedia learning, Visual literacy, Science education

How to cite:
Loke, W. M. (2026). Visual Design in High-Stakes Science Assessments: A Longitudinal Content Analysis of Singapore Primary Science Examinations. *Journal of Science Education Research*, 10(2), 1-9. doi: https://doi.org/10.21831/jser.v10i2.101105.

1. INTRODUCTION

Contemporary science education is fundamentally multimodal, requiring learners to seamlessly transition between verbal, mathematical, and visual modes to develop scientific understanding (Padios et al., 2023; Unsworth et al., 2022). Visual representations, ranging from concrete photographs to highly abstract schematic systems, are not merely supplementary elements but are fundamental to the epistemic practice of science (Kress & van Leeuwen, 2021). Developing what Ainsworth (2021) defines as 'representational competence', which describes the capacity to interpret, translate, and transform diverse representational modes appropriately, is therefore a foundational objective of primary and secondary science curricula globally. Visuals act as vital cognitive bridges, anchoring abstract conceptual frameworks within observable, spatialized forms, especially in early science education (Justi & Gilbert, 2016). Therefore, the ability to interpret visual information is just as crucial to understanding science as the skill of reading text (Fitri & Rahmawati, 2018).

Textbooks and lectures leverage visuals effectively, yet summative exams impose a distinct cognitive regime. Sweller et al. (2019) define working memory as strictly finite, forcing items to balance intrinsic task demand against presentation format (Ehrich et al., 2021). Visuals act as cognitive scaffolds, clarifying text, lowering linguistic thresholds, and anchoring situational models (Mayer, 2024; Rexigel et al., 2024). Mayer's Coherence Principle supplies a prescriptive boundary: exclude any element lacking structural necessity. Poorly designed or extraneous visuals can induce split-attention effects or introduce ‘seductive details’ that misdirect a student’s limited working memory capacity away from the core construct being evaluated (Sundararajan & Adesope, 2020; Bender et al., 2021). Evaluating how visual elements are structured within assessment interfaces is therefore essential to determine whether item designs minimize unnecessary visual friction while maintaining appropriate task complexity. The author applied these cognitive frameworks to a national standardized testing ecosystem to systematically analyze the structural and functional composition of graphics embedded in exam items. Prior research in science education has increasingly used visual taxonomies to analyze instructional texts and classroom tasks (Ate et al., 2025; Tytler et al., 2018). Empirical examinations characterizing how visual typologies and functional necessity interact within high-stakes primary summative instruments remain limited.

Nowhere is this assessment design imperative sharper than in centralized, high-performing systems like Singapore, where the Primary School Leaving Examination (PSLE) functions as the critical summative gateway (Ministry of Education, Singapore, 2021; Mullis et al., 2021). The high-stakes nature of national examinations generates a documented washback effect, shaping visual-literacy priorities within primary science classrooms (Lim & Tan, 2021). Consequently, the visual genres and interpretive competencies prioritized in summative assessments strongly influence the semiotic skills emphasized during classroom instruction. Visual representations in science education operate across a spectrum of domain-specific semiotic systems (Kress & van Leeuwen, 2021; Padios et al., 2023). Prior research partitions visual typography into four major categories: (a) Representational graphics showing physical isomorphism to real-world objects; (b) Abstract and Symbolic diagrams utilizing domain-specific notation codes (e.g., circuit schematics, vector fields); (c) Data-Driven representations encoding quantitative relationships (e.g., graphs, data tables); and (d) Relational schematics mapping process flows and biological cycles (Ate et al., 2025; Tytler et al., 2018). In parallel, cognitive architectures dictate the functional utility of graphics within task environments, distinguishing between essential visual inputs required for item solution, supporting scaffolds that clarify verbal prompts, and decorative details that add non-diagnostic visual noise (Mayer, 2024; Rexigel et al., 2024). To evaluate this visual landscape systematically, the study operationalized a dual-axis taxonomy adapted from established visual semiotic frameworks in science education literature (Ate et al., 2025; Coleman & Dantzler, 2016; Tytler et al., 2018). The first axis captured formal typography (representational, abstract/symbolic, data-driven, and relational, while the second axis captured functional purpose (essential, supportive, and decorative) (Mayer, 2024; Tytler et al., 2018). While prior studies have examined visual representations in science trade books (Coleman & Dantzler, 2016), digital learning media (Fitri & Rahmawati, 2018; Wicaksono & Kuswanto, 2026), and instructional tasks (Ate et al., 2025; Padios et al., 2023), few studies have systematically tracked how visual categories and functional roles shift across question formats and content domains over time within national examination papers. Rather than proposing an entirely new taxonomy, the novelty of this study lies in combining these established typographical and functional frameworks to perform a total-population, longitudinal content analysis (2023-2025) of visual design patterns within a national primary science examination corpus.

To address this empirical gap, the present study conducts a comprehensive total-population content analysis of the visual elements embedded within the Singapore PSLE Science examinations across a three-year longitudinal window (2023–2025). Rather than using a sampled corpus, this study evaluates every visual representation used in both the Multiple-Choice Question (MCQ) and Short-Answer Question (SAQ) components of the national papers. The investigation is guided by three primary research objectives:

1. To characterize the typological distribution of visual representations across the five core thematic blocks of the national primary science syllabus (Cycles, Diversity, Energy, Interactions, and Systems).
2. To evaluate the functional necessity of these visuals across test item stems .
3. To analyze structural variations in visual configurations across question formats (MCQ vs. SAQ) and track longitudinal shifts in visual deployment from 2023 to 2025.

By interrogating these patterns, this study establishes an empirical baseline for understanding how visual literacy is assessed at the culmination of primary education, offering critical insights for curriculum developers, item writers, and science educators internationally.

2. RESEARCH METHOD

Research Design

This study utilized a quantitative, longitudinal content analysis design (Krippendorff, 2019) to systematically evaluate visual representations in national primary science examination papers. A deductive coding framework was established to isolate structural typologies, functional utility, and longitudinal trends across question formats (MCQ vs. SAQ). The primary unit of analysis was defined as any discrete graphical element—including photographs, schematics, charts, data tables, and flowcharts—embedded within an item stem or choice option. Where multiple discrete graphics formed an integrated conceptual setup (e.g., an experimental set-up paired with a key), the composite graphic was coded as a single perceptual visual unit to preserve the item's holistic cognitive demand.

Data Source and Corpus

The analytical corpus comprised the complete population of official national Primary School Leaving Examination (PSLE) Standard Science examination papers administered in Singapore across three consecutive annual cycles: 2023, 2024, and 2025 (Singapore Examinations and Assessment Board, 2023–2025). All examination papers were retrieved from officially published, SEAB-authorized past-year national examination compilations published under license by approved educational distributors.

To ensure structural equivalence and comparability across the longitudinal window, the corpus was strictly confined to the standard main-sitting operational examination papers (Standard Science syllabus, Subject Code 0009). Special variant sittings, supplementary papers, and alternative curriculum tracks (such as Foundation Science papers) were excluded from the dataset. The structure of the standard national examination paper is standardized by the Singapore Examinations and Assessment Board (SEAB) and the Ministry of Education (MOE), Singapore. Each annual paper consists of Section A (28 Multiple-Choice Questions, 2 marks each) and Section B (12 Short-Answer Questions, 2 to 5 marks each), yielding exactly 40 items per annum and a cumulative three-year population corpus of 120 items ($N_{MCQ} = 84$; $N_{SAQ} = 36$).

Regarding copyright and fair dealing provisions, the examination materials were analyzed strictly for academic content analysis under standard non-commercial research principles. No proprietary test items, copyrighted illustrations, or full-page scan outputs are reproduced verbatim within this manuscript. All data are presented as aggregated statistical metrics, thematic code counts, and non-proprietary structural taxonomies.

To fulfill the first and third research objectives, every extracted visual element and its parent test item were stratified across the five official thematic blocks mandated by the national primary science curriculum framework: Cycles, Diversity, Energy, Interactions, and Systems.

Visual Classification Framework (The Taxonomy)

The primary analytical instrument was a dual-axis visual coding rubric adapted from established visual semiotic and cognitive frameworks in science education (Ate et al., 2025; Mayer, 2024; Tytler et al., 2018). The instrument evaluates graphics across two dimensions:

- Axis 1: Formal Typology (Morphology)
  - Representational: Concrete drawings or photographs displaying physical isomorphism to macroscopic or microscopic real-world entities (e.g., lab apparatus setups, plant anatomy).
  - Abstract and Symbolic: Non-iconic diagrams utilizing arbitrary symbols and disciplinary codes requiring specialized domain knowledge (e.g., IEEE circuit schematics, vector force arrows).
  - Data-driven: Graphical structures encoding quantitative relationships or discrete measurement matrices (e.g., line graphs, bar charts, numeric data tables).
  - Relational: Non-numeric schematics mapping process dynamics, conceptual networks, or sequential phases (e.g., food webs, life cycle flowcharts).
- Axis 2: Functional Utility (Cognitive Purpose)
  - Essential: Visuals functioning as indispensable primary data carriers required to solve the task.

- Supporting: Visuals providing redundant or parallel spatial clarifying scaffolds alongside explicit verbal text.
- Decorative: Visuals introduced purely for aesthetic appeal, carrying no task-relevant data (Sundararajan & Adesope, 2020).

Coding Procedure and Reliability

Face and content validity of the coding instrument were established by a panel of two educational researchers specializing in primary science assessment design and visual semiotics. To establish operational reliability, raters conducted a pilot coding session on 15 visual units from pre-2023 PSLE papers to standardize boundary definitions. Using the finalized instrument, both raters independently evaluated the 2023–2025 corpus (N = 120 items; N = 134 visuals) in a double-blind format (O'Connor & Joffe, 2020). Inter-rater reliability was calculated using Cohen’s kappa to account for chance agreement (Ten Hove et al., 2024). The initial coding yielded kappa = 0.89 for typological classification and kappa = 0.92 for functional classification, indicating "almost perfect" agreement. Discrepancies were resolved through a joint reconciliation review until 100% consensus was achieved.

Data Analysis

Data matrices were managed and formatted using Microsoft Excel (ver. 2604, build 19929.20164), while all statistical analyses were executed using IBM SPSS Statistics (Version 29.0). Analytical techniques were structured directly to address the three core research objectives:

Objective 1 (Typological Distribution): Evaluated using absolute frequencies (n), percentage distributions (Freq. %), and Pearson’s Chi-Square Test of Independence ($\chi 2$) to test dependencies between visual typologies and scientific themes (Cycles, Diversity, Energy, Interactions, Systems) (Agresti, 2018).

Objective 2 (Functional Necessity): Evaluated via cross-tabulation of functional tiers (Essential, Supporting, Decorative) against visual typologies and assessment components.

Objective 3 (Format Variations and Longitudinal Trends): Evaluated using Pearson’s Chi-Square Test of Independence ($\chi 2$) to compare typological distributions between question formats (MCQ vs. SAQ). Fisher’s Exact Test was substituted whenever expected cell frequencies fell below 5. Annual subtotals (2023–2025) were tracked to analyze longitudinal shifts. All inferential tests applied a two-tailed significance threshold of $p < 0.05$.

3. RESULT AND DISCUSSION

The empirical deconstruction of the Singapore PSLE Science papers (2023–2025) provides an observable benchmark of visual representation design within standardized national examinations. Across the three-year corpus (N = 120 test items), a total of 134 discrete visual elements were identified and evaluated across content domains, functional utility tiers, and item formats.

Typological Distribution of Visual Representations Across Core Themes

To address the first research objective, extracted visual elements were classified into four structural typologies across the five official primary science thematic blocks (Table 1). Representational visuals constituted the largest share of the corpus (n = 57, 42.54%), followed by Data-Driven visuals (n = 44, 32.84%) and Abstract & Symbolic graphics (n = 29, 21.64%). Relational visuals comprised a highly specialized fraction of the total dataset (n = 4, 2.99%).

The volume of visuals varied across curriculum themes: Interactions contained the highest visual density (n = 43, 32.09%), followed by Systems (n = 29, 21.64%), Cycles (n = 28, 20.90%), Energy (n = 21, 15.67%), and Diversity (n = 13, 9.70%).

Table 1. Typological Distribution of Visual Elements Across Scientific Themes (2023–2025)

Scientific Theme	Representational	Abstract & Symbolic	Data-Driven	Relational	Total
Cycles	15	2	7	4	28
Diversity	7	1	5	0	13
Energy	7	4	10	0	21
Interactions	16	12	15	0	43

Scientific Theme	Representational	Abstract & Symbolic	Data-Driven	Relational	Total
Systems	12	10	7	0	29
Overall Total	57	29	44	4	134

A Pearson's Chi-Square Test of Independence revealed a statistically significant relationship between visual typologies and scientific themes ($\chi2(12) = 26.76$, $p = 0.0084$). This indicates that visual deployment varies systematically according to topic domain. The Cycle theme accounted for 100% of all Relational process visuals (n = 4) in the corpus, reflecting the domain's focus on cyclical biological and physical transformations (e.g. life cycles and water cycle). The Cycle theme also maintained a strong baseline of Representational graphics (n = 15). The Interactions and Systems themes exhibited the higher proportions of Abstract & Symbolic visuals (n = 22, 75.86%), such as electrical circuit schematics and force vector arrows. These distributional patterns align with visual semiotics literature in science education (Ainsworth, 2021; Ate et al., 2025; Tytler et al., 2018), which notes that physical-centric domains rely heavily on abstract symbolic codes that do not physically resemble their real-world referents, requiring domain-specific visual decoding skills. Visual deployment within the Energy theme (n = 21) was led primarily by Data-Driven matrices (n = 10, 47.62%) and Representational models (n = 7, 33.33%). Data-Driven matrices present co-varying quantitative variables for empirical extraction, whereas Representational models depict physical apparatus setups to spatially anchor unobservable energy transformations. Pairing concrete structural diagrams with empirical data tables operationalizes dual-coding architecture (Hansen & Richland, 2020; Mayer, 2024; Sweller et al., 2019). This visual pairing reduces extraneous linguistic processing load by establishing direct perceptual mapping between physical test variables and their corresponding quantitative outputs. Visual deployment in the Diversity domain is dominated by Representational drawings (53.85%, n = 7) and Data-Driven tables (38.46%, n = 5), reflecting an empirical focus on observational classification and attribute comparison.

Functional Necessity of Visual Elements in High-Stakes Assessments

Regarding the second research objective, visual elements were categorized by operational utility within the task-solving path. Across the 134 visual units, 97.01% (n = 130) were categorized as functionally Essential, 2.99% (n = 4) served a Supporting role, and 0.00% (n = 0) were classified as Decorative. All Abstract & Symbolic (n = 29), Data-Driven (n = 44), and Relational (n = 4) visuals were functionally essential; item prompts were incomplete without them. Non-essential visuals were confined to four Representational drawings (7.02% of representational graphics) that provided redundant spatial clarifications alongside explicit verbal text.

The complete absence of decorative graphics (0.00%) reflects an alignment with Mayer's (2024) Coherence Principle as an analytical benchmark. While instructional materials often include decorative illustrations to stimulate situational interest, the examined national examination papers excluded non-diagnostic artwork. This design pattern avoids introducing extraneous perceptual details that could draw attention away from task-relevant information. Supporting visuals comprised a minor share of the overall corpus (2.99%, n = 4) and were restricted exclusively to representational graphics. Rather than serving as decorative artwork, these redundant visual elements functioned as concrete spatial scaffolds to clarify verbal prompts and lower linguistic processing friction (Mayer, 2024; Schnotz, 2021).

Comparative Analysis of Visual Deployment by Question Format

To address the third research objective, visual distributions were compared across item formats. MCQs contained 64.93% (n = 87) of all visual elements, while SAQs contained 35.07% (n = 47). A Pearson's Chi-Square Test of Independence indicated no statistically significant difference in the overall proportional makeup of visual typologies between formats ($\chi2(3) = 3.91$, $p = 0.2715$), indicating that the structural distribution of visual typologies remains consistent across both item formats.

Despite overall proportional consistency, absolute distributions revealed descriptive differences. MCQs contained higher total counts of Data-Driven visuals (n = 31, 70.45%) and Abstract & Symbolic diagrams (n = 20, 68.97%), using graphics as data extraction sources from which examinees select correct option choices. In contrast, SAQs exhibited a higher relative proportion of Representational graphics (46.81% of SAQ visuals; n = 22 vs. 40.23% in MCQs; n = 35). In open-ended items, concrete representational visuals provide situational context alongside written response prompts (Ainsworth, 2021; Mayer, 2024). Rather than revealing functional distinctions or format-specific assessment strategies, these distributions reflect a structurally balanced demand for data interpretation and visual literacy across both objective multiple-choice and open-ended question formats. Descriptively, visual deployment across item formats reveals distinct structural patterns. In MCQs, Data-Driven (n = 31) and Abstract & Symbolic (n = 20) graphics predominate, acting as visual data sources from which examinees extract information to evaluate option choices. Conversely, SAQs feature a higher relative proportion of Representational drawings (46.81%, n = 22), providing concrete situational context alongside open-ended written prompts (Ainsworth, 2021; Mayer, 2024).

Longitudinal Trends in Visual Integration (2023–2025)

Tracking the corpus longitudinally revealed an overall contraction in total visual elements, declining from 54 in 2023, to 46 in 2024, and to 34 in 2025. Decoupling the data by question format revealed contrasting trajectories (Figure 1). Visual elements in MCQs experienced a continuous reduction, falling from 39 (2023) to 27 (2024) and 19 (2025). This was driven by declines in Data-Driven graphics (14 in 2023 down to 8 in 2025) and Abstract elements (8 in 2023 down to 3 in 2025). Visual elements in SAQs remained stable (15 in 2023, 19 in 2024, 15 in 2025). Within SAQs, Representational visuals nearly doubled over three years, rising from 6 (2023) to 10 (2024) and 11 (2025).

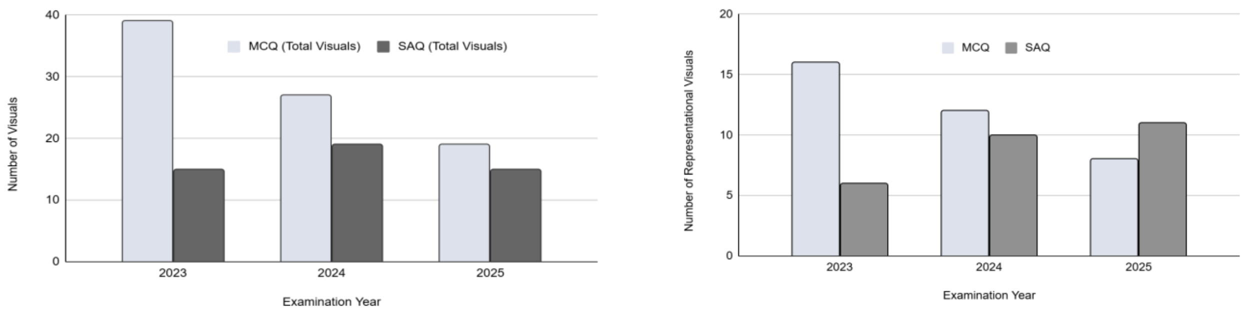

Figure 1. Longitudinal Shift in Visual Deployment by Question Format (2023–2025).

Longitudinal data from 2023 to 2025 highlighted a distributional shift in visual elements across item formats. Visuals embedded within the MCQ section decreased from 39 to 19 elements, whereas the visual footprint within the SAQ section remained stable, accompanied by an increase in representational drawings (n = 6 in 2023 to n = 11 in 2025). Descriptively, this trajectory indicates a structural reallocation across formats: representational imagery became less prevalent in objective selection tasks (MCQs) and more prominent within open-ended items (SAQs). Within SAQs, concrete representational graphics provide situational context alongside written prompts, aligning with multimedia learning frameworks on how spatial anchors support mental model construction during explanatory tasks (Fitri & Rahmawati, 2018; Noetel et al., 2022; Schnotz, 2021). Descriptively, this trajectory marks a pattern where visual elements in objective questions function primarily as data sources, while those in open-ended questions function as contextual scaffolds (Tytler et al., 2018). This shift reflects a descriptive pattern where graphics in open-ended items serve as contextual anchors for written responses rather than complex data puzzles in multiple-choice stems.

Implications for Science Assessment Design and Pedagogy

These observational findings offer practical insights for curriculum developers, test creators, and educators. The high proportion of essential graphics (97.01%) and total exclusion of decorative visuals (0.00%) demonstrate a strict baseline of functional necessity. Item developers should maintain this standard by ensuring graphics serve a direct diagnostic role within the item stem, eliminating non-essential artwork that could introduce extraneous cognitive processing demands (Mayer, 2024; Sundararajan & Adesope, 2020). The significant association between visual typologies and subject themes highlights the need for domain-specific visual literacy instruction. Because abstract symbols dominate Systems and Interactions while relational flowcharts cluster in Cycles, teachers must explicitly teach students how to decode specialized notation systems (e.g., circuit schematics, vector arrows, process cycles) as formal disciplinary languages (Ainsworth, 2021; Ate et al., 2025). As representational visuals increase in open-ended items (SAQs), instruction should focus on helping students use situational visual anchors to structure written explanations, bridging the transition between visual interpretation and text generation.

Limitations of the Study

While this study provides a total-population baseline of visual representations within a national primary science examination, several methodological boundaries should be acknowledged. The dataset is restricted to a three-year longitudinal window (2023–2025) comprising 120 test items (N = 134 visuals). While analyzing a complete three-year population corpus eliminates sampling bias, tracking visual design trajectories across a wider multi-year window would be necessary to establish broader, long-term trends in assessment policy. The investigation focused exclusively on the Singapore PSLE Standard Science syllabus. Because national assessment ecosystems vary in language profiles, layout conventions, and syllabus structures, these findings may not directly generalize to decentralized examination boards, secondary-level science assessments, or different international systems. This study evaluated the structural design and functional roles of visual elements from an assessment construction perspective. While findings were interpreted through Cognitive Load Theory and Mayer's Coherence Principle as analytical frameworks, the research did not directly measure student-level cognitive processing, eye-tracking data, or psychometric item difficulty metrics. The corpus excluded Foundation Science papers (Subject Code 0039) and specimen mock papers to maintain cohort comparability. Future empirical research should incorporate eye-tracking methodologies and psychometric performance data to evaluate how students actively process these visual typologies during task execution. Additionally, comparative studies examining Foundation Science (Subject Code 0039) or international assessment frameworks could reveal how visual design strategies are adapted across diverse learner cohorts and testing environments.

4. CONCLUSIONS

This study provides a total-population empirical benchmark of visual design patterns within a national primary science examination framework. The findings demonstrate that visual deployment is not uniform across testing materials; rather, assessment design relies on domain-specialized graphical conventions that systematically align specific syllabus topics with distinct visual typologies. Furthermore, the functional essentiality of visual elements reflects an item-writing baseline aligned with Mayer's Coherence Principle as an analytical framework, where visuals serve as diagnostic data sources rather than decorative embellishments. The observed longitudinal shift across item formats—streamlining quantitative visual extraction in multiple-choice questions while expanding concrete visual anchors in short-answer questions—highlights an evolving balance between objective selection and open-ended response tasks. These empirical patterns carry direct implications for both assessment practices and science pedagogy. For item developers, maintaining strict functional necessity preserves construct validity while preventing extraneous cognitive processing friction. For educators, the domain-specific nature of visual deployment underscores that visual literacy cannot be treated as a generic skill. Science instruction must explicitly scaffold the decoding of subject-tailored semiotic systems, ranging from abstract symbolic notations in physical systems to sequential relational flowcharts in cyclic processes.

REFERENCES

Agresti, A. (2018). *An introduction to categorical data analysis* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/0470114754

Ainsworth, S. (2021). The multiple representations principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 158–170). Cambridge University Press. https://doi.org/10.1017/9781108894333.016

Ate, O., Parno, P., & Munfaridah, N. (2025). Multirepresentation in physics learning research: A content analysis review of trends, opportunities, and challenges. *Journal of Science Education Research*, 9(1), 33–41. https://doi.org/10.21831/jser.v9i1.72954

Bender, L., Renkl, A., & Eitel, A. (2021). Seductive details do their damage also in longer learning sessions—When the details are perceived as relevant. *Journal of Computer Assisted Learning*, 37(5), 1248–1262. https://doi.org/10.1111/jcal.12560

Coleman, J. M., & Dantzler, J. A. (2016). The frequency and type of graphical representations in science trade books for children. *Journal of Visual Literacy,* 35(1), 24–41. https://doi.org/10.1080/1051144X.2016.1198543

Ehrich, J. F., Howard, S. J., Bokosmaty, S., & Woodcock, S. (2021). An item response modeling approach to cognitive load measurement. *Frontiers in Education*, 6, 648324. https://doi.org/10.3389/feduc.2021.648324

Fitri, F., & Rahmawati, L. (2018). Development of Kinect-based science learning media to support improvement of cognitive learning outcomes. *Journal of Science Education Research*, 2(1), 9–12. https://doi.org/10.21831/jser.v2i1.19261

Hansen, J., & Richland, L. E. (2020). Teaching and learning science through multiple representations: Intuitions and executive functions. *CBE—Life Sciences Education*, 19(4), ar61. https://doi.org/10.1187/cbe.19-11-0253

Justi, R., & Gilbert, J. K. (2016). *Modelling-based teaching in science education*. Springer. https://doi.org/10.1007/978-3-319-29039-3

Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge.

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications. https://doi.org/10.4135/9781071878781

Lim, F. V., & Tan, J. M. (2021). Curriculum and assessment mismatch: Examining the role of images in literacy assessments. *The Australian Journal of Language and Literacy*, 44(3), 22–34. https://doi.org/10.1007/BF03652078

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, 36(1), 8. https://doi.org/10.1007/s10648-023-09842-1

Ministry of Education, Singapore. (2021). *Science syllabus: Primary 3 to 6*. Curriculum Planning and Development Division.

Mullis, I. V., Martin, M. O., & von Davier, M. (2021). *TIMSS 2023 assessment frameworks: 28 years of trends*. International Association for the Evaluation of Educational Achievement.

Noetel, M., Griffith, S., Delaney, O., Harris, N. R., Sanders, T., Parker, P., Delaney, J., & Lonsdale, C. (2022). Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Review of Educational Research*, 92(3), 413–454. https://doi.org/10.3102/00346543211052329

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1–13. https://doi.org/10.1177/1609406919899220

Padios, A., Pascua, S., & Orleans, A. (2023). Multimodal representations across scientific genres: Perspectives from science teachers' practices. *International Journal of Education*, 16(2), 89–100. https://doi.org/10.17509/ije.v16i2.55412

Rexigel, E., Kuhn, J., Becker, S., & Malone, S. (2024). The more the better? A systematic review and meta-analysis of the benefits of more than two external representations in STEM education. *Educational Psychology Review*, 36(4), 124. https://doi.org/10.1007/s10648-024-09958-y

Schnotz, W. (2021). Integrated model of text and picture comprehension. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 82–99). Cambridge University Press. https://doi.org/10.1017/9781108894333.010

Singapore Examinations and Assessment Board. (2023–2025). *Singapore Primary School Leaving Examination (PSLE) Science examination paper*. Ministry of Education, Singapore.

Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, 32(3), 707–734. https://doi.org/10.1007/s10648-020-09522-4

Sweller, J., Van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Ten Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2024). Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological Methods*, 29(5), 967–985. https://doi.org/10.1037/met0000516

Tytler, R., Prain, V., & Hubber, P. (2018). Representation construction as a core science disciplinary literacy. In D. A. Treagust, R. Duit, & H. E. Fischer (Eds.), *Global developments in literacy research for science education* (pp. 301–317). Springer. https://doi.org/10.1007/978-3-319-69197-8_18

Unsworth, L., Tytler, R., Fenwick, L., Humphrey, S., Chandler, P., Herrington, M., & Pham, L. (2022). *Multimodal literacy in school science: Transdisciplinary perspectives on theory, research and pedagogy*. Routledge. https://doi.org/10.4324/9781003150718

Wicaksono, B. A., & Kuswanto, H. (2026). Enhancing senior high school students' critical thinking and graphical representation skills through an AR-integrated guided inquiry e-book. *Journal of Science Education Research*, 10(1), 101–112. https://doi.org/10.21831/jser.v10i1.81024