JOURNAL OF MATERIALS EDUCATION (Electronic)

Original Research

Selecting Age-Dependent Concrete Strength Comparisons for Offline Materials Instruction

Daniel Ocen Okello[1], Grace Lamunu Akello[2], Samuel Odong Opio[3], Esther Adoch Ojara[3], Patrick Komakech Otema[4], Miriam Acen Olara[1]

[1] *Department of Biosystems Engineering, Faculty of Agriculture and Environment, Gulu University, Gulu, Uganda*
[2] *Department of Science and Vocational Education, Faculty of Education, Lira University, Lira, Uganda*
[3] *Department of Curriculum and Instruction, Faculty of Education, Muni University, Arua, Uganda*
[4] *Department of Curriculum and Educational Technology, Faculty of Education and Humanities, Gulu University, Gulu, Uganda*

Correspondence to:
daniel.ocen@gu.ac.ug

RECEIVED:
02/05/2026
ACCEPTED:
21/09/2026
AVAILABLE ONLINE:
31/10/2026

ABSTRACT

Materials instruction must distinguish a measured strength ordering from a claim that remains valid at another specimen age. This investigation develops an offline activity that records a prediction before revealing later-age compressive strengths and evaluates the selection of comparisons without repeating formulations. The principal analysis uses 115 formulations measured at both seven and twenty-eight days. Requiring a strength separation of at least 3 MPa at each age leaves 5,141 admissible comparisons, including 185 reversed orderings. Reversal-prioritized selection supplies ten examples involving twenty different formulations. Among a thousand randomly ordered ten-case sessions, 732 contain no reversal, whereas selection by the largest early-strength difference exposes none. Integer optimization establishes that the principal comparison pool can support thirty formulation-disjoint reversals. Selection-order losses become apparent at a stricter seven-to-ninety-day comparison: greedy selection delivers six reversals, although nine are attainable. Alternative age intervals, contrast margins and treatments of repeated measurements distinguish numerical sensitivity from limitations in case supply. Seven graphical analyses characterize observation coverage, shared-formulation dependence, attainable selection, random delivery, contrast magnitude, rank movement and simultaneous composition changes. The instructional contribution is a precise connection between measurement conditions, comparison selection and staged evidence disclosure.

KEYWORDS

materials instruction, concrete strength, specimen age, offline delivery, rank reversal, formulation exclusion

© 2026 The Authors

This is an open access article published by InterPore under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0) (https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. INTRODUCTION

Interpreting a material property requires attention to the conditions attached to its measurement. A concrete mixture can exceed another in compressive strength at one age and fall below it at another, even when both mixtures become stronger over time. Students who interpret strength as an unqualified label may overlook this distinction. The resulting error is more subtle than an incorrect subtraction: the arithmetic can be accurate while

the inference extends beyond the conditions supporting it. Teaching therefore needs comparisons that make the boundary of a claim visible. Contemporary discussion of concrete education places such decisions alongside the practical demands of preparation time, available equipment and classroom organization [1].

Age dependence has a physical basis in the evolution of cement hydration and the development of load-bearing microstructure. Reaction rates, the balance of hydrate phases and the connectivity of pore space influence strength development rather than producing a single invariant relationship between ingredient quantities and mechanical response [2]. Supplementary cementitious constituents further complicate comparisons because their chemical composition, fineness and reactivity vary, with consequences extending beyond an early test [3]. Students need to recognize these possibilities without treating an observational comparison as proof of a particular reaction mechanism. A useful teaching case can demonstrate that an age-independent assertion fails while leaving the mechanistic explanation open for additional evidence.

The distinction has practical relevance when discussing concrete with reduced clinker content. Performance requirements must be considered alongside environmental objectives, availability of replacement constituents and the time at which mechanical properties are needed [4, 5]. An activity that displays only a favorable early-strength ranking can encourage a premature preference for one composition. Conversely, showing a later reversal without explaining its evidential limits can encourage the equally weak rule that a mixture with lower initial strength will eventually become superior. Instruction should therefore support a conditional statement tied to the measured ages, rather than replace one universal shortcut with another. No environmental inventory is calculated here; the sustainability literature explains why careful comparison matters, not which of the studied mixtures is environmentally preferable.

A useful distinction for students is between describing a specimen, comparing two recipes and recommending a material for an application. Each activity requires different evidence. A specimen description identifies the measured response and its test conditions. A comparison adds another material and requires compatible quantities. A recommendation introduces a service requirement, which may involve construction timing, durability, cost or environmental consequences that a strength table does not contain. The present activity addresses the middle task. Making that scope explicit allows an instructor to discuss the relevance of a reversed strength ordering without implying that the later winner is the preferred construction material. It also supplies a practical reason for keeping measurement units and age together throughout the interaction: both are necessary to state the comparison before a broader engineering judgment can be justified.

Digital delivery offers a way to organize the encounter with such evidence. Virtual compression-testing activities have already been used to connect theory with laboratory practice in large engineering classes [6]. Materials-education research has also considered the assembly and sequencing of reusable learning objects [7]. These contributions establish that neither browser delivery nor ordering instructional content is itself a new educational technique. The narrower issue is how an instructor can select measured comparisons that challenge an age-independent claim and deliver those comparisons while preserving the distinction between an initial prediction and the later observation.

Reviews of virtual laboratories show that technology, instructional support and the assessed outcome must be considered together. Evidence spanning undergraduate science and engineering does not support treating every digital activity as equivalent to every laboratory experience [8]. Research on virtual chemistry laboratories similarly separates the educational possibilities of digital environments from the need to examine practical skills and the support provided during an activity [9]. The present resource does not attempt to teach specimen casting, machine operation or safe testing practice. It concentrates on interpreting a strength comparison after the age condition changes, which is a smaller and more directly inspectable instructional task.

Contrasting cases offer an appropriate starting point because meaningful variation can draw attention to relations that students would otherwise overlook. In engineering activities, the recognition of underlying scientific relations has been associated with performance and transfer to other situations [10]. Evidence also indicates that an exploratory activity containing contrasting cases can differ educationally from an activity that merely presents many observations [11]. The distinction is important for materials teaching: a large table may contain useful counterex-

amples without ensuring that a short classroom session actually exposes one. The selection of a few comparisons is therefore part of the instructional design, not simply a convenient reduction in the amount of information displayed.

The timing of an explanation remains consequential. Research on problem solving followed by instruction identifies conditions under which preliminary exploration can prepare students to learn from a subsequent explanation [12]. A broader analysis of this literature indicates that the advantages depend on the learners, the targeted knowledge and the design of the learning sequence [13]. Direct examination of contrasting cases before and after instruction further cautions against assuming that an opportunity to compare will automatically produce better understanding [14]. The activity developed here consequently uses deferred evidence as a testable delivery decision. Its educational rationale does not substitute for an evaluation with students.

Prediction provides a specific action that can precede disclosure. Asking learners to commit to an expected outcome can create an opportunity to compare that expectation with what subsequently becomes visible [15]. Experiments using numerical facts have found that explicit predictions can benefit memory for outcomes that violate expectations [16]. Such findings support investigating a prediction-and-reveal interaction, while leaving open whether the same benefit occurs when learners reason about cementitious materials. A further distinction is needed between making a prediction and asserting that the existing evidence determines the answer. An "undetermined" response is scientifically legitimate when the early-age observations do not uniquely establish the later ordering.

The effectiveness of written responses also depends on what students are asked to generate. Research on explaining, visualizing and acting on learning material emphasizes the quality of the relations learners construct, rather than the mere presence of an activity [17]. Comparisons between pretesting and posttesting likewise show that the format of an initial attempt and the feedback accompanying it matter [18]. The resource therefore asks for a reason before disclosure and an age-qualified statement afterwards. A minimal character requirement ensures that text is entered; it cannot determine whether that text contains a scientifically defensible explanation. This distinction prevents a software completion check from becoming an unsupported judgment about understanding.

The broader learning literature strengthens the need for restraint when interpreting participation. Active engagement can produce learning experiences that feel more difficult even when measured learning improves [19]. Evidence concerning active learning and participation across undergraduate science and engineering also shows why instructional arrangements should be evaluated with attention to who benefits [20]. Neither enjoyment nor the number of completed cases is measured in this investigation. The numerical evaluation instead asks whether a particular type of material evidence is made available under a fixed session length. Any later educational comparison must assess the quality of reasoning independently of interface completion.

Concrete-strength prediction has advanced through statistical analysis, ensemble learning and interpretable methods [21–23]. Reviews document a much wider range of predictive approaches across concrete types and properties [24]. That literature addresses estimation of an unknown response from inputs. Selecting instructional comparisons is a different operation because the instructor can inspect both measured ages before choosing a case. The analysis here deliberately uses that access. It evaluates the availability and arrangement of known contrasts, without presenting the resulting selection advantage as predictive accuracy or evidence that a learner could anticipate the later strengths.

A remaining difficulty is dependence between comparisons. A striking formulation can participate in many reversed pairs, so a high pair count does not necessarily support a diverse set of teaching examples. Counting comparisons alone may conceal a delivery sequence dominated by the same material. This investigation treats each formulation as available only once within a session and examines the consequences through both greedy selection and exact integer optimization. The contribution is an offline instructional resource accompanied by a quantitative account of which age-conditioned comparisons can be delivered, which selection losses can be avoided, and which claims remain unsupported by the available measurements. The analysis concerns existing laboratory observations and new calculations; it includes no newly cast specimens or measured student outcomes.

2. MEASUREMENTS AND INSTRUCTIONAL METHOD

2.1. Material Quantities, Age Coverage and Units

A recent study investigated the relationship between concrete mixture composition, curing age, and compressive strength, providing valuable insights into the influence of constituent materials on mechanical performance [25, 26]. The investigation considered 1,030 concrete mixtures characterized by seven material quantities, including cement, blast-furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate, together with specimen age and compressive strength. Material quantities were expressed in kilograms per cubic metre, curing age in days, and compressive strength in megapascals. The research established a useful foundation for examining how variations in mixture proportions and curing duration relate to concrete strength development. These findings also provide a practical basis for comparative analysis, quantitative interpretation, and instructional activities focused on composition–property relationships in cementitious materials.

Exact equality across all nine quantities identifies twenty-five repeated rows. Removing these leaves 1,005 distinct observations. Formulations are then identified through exact equality of the seven ingredient quantities, giving 428 combinations. Lexicographic ordering of those combinations assigns the internal identifiers M001 to M428. These labels identify numerical recipes for comparison; they do not identify verified experimental batches. The ingredient and strength ranges appear in Table 1. Keeping the stated units visible in the activity prevents a normalized input value or an ingredient fraction from being mistaken for a measured strength.

Table 1: Quantities for material comparison.

Quantity	Unit	Range
Cement	kg/m^3	102.00–540.00
Blast-furnace slag	kg/m^3	0.00–359.40
Fly ash	kg/m^3	0.00–200.10
Water	kg/m^3	121.75–247.00
Superplasticizer	kg/m^3	0.00–32.20
Coarse aggregate	kg/m^3	801.00–1,145.00
Fine aggregate	kg/m^3	594.00–992.60
Specimen age	day	1–365
Compressive strength	MPa	2.33–82.60

Grouping by formulation and age produces 996 cells. Nine contain two different strengths, while the other cells contain one strength each. Their median provides a descriptive cell value without assigning a standard error. Records at ninety and ninety-one days remain distinct, and unobserved ages are not interpolated. Figure 1 shows that the ages are unevenly represented. A formulation enters an age comparison only when both required cells are present, so a large count at twenty-eight days cannot compensate for an absent earlier or later observation. The principal seven-to-twenty-eight-day comparison contains 115 formulations, enough to construct a ten-case session using twenty distinct recipes.

The distinction between recipe equality and specimen identity also governs the interpretation of repeated values. Removing an exact repeated row prevents it from increasing the apparent amount of evidence, but it does not establish whether the repeated entry represents a copying error or a repeated physical test. Similarly, taking a median of two different strengths gives the activity a single numerical value without identifying the cause of their difference. The analysis therefore retains the number of entries associated with each cell when checking the selected cases. This treatment preserves the observation structure while avoiding an unsupported variance estimate. An independently replicated material experiment would require specimen and batch information that is unavailable for these comparisons.

The age and ingredient quantities are the conditions explicitly retained in the analysis. Curing temperature, humidity, specimen geometry, testing standard, laboratory identity and independent batch labels are unavailable in the tabulation. Consequently, matching a recipe across ages does not verify that every other experimental condition was held constant. Distinct recipes provide distinct computational case units, but the number of independent

physical samples cannot be established. This boundary is sufficient for a descriptive teaching-resource question and insufficient for estimating a controlled effect of age or a particular constituent. It is carried into the interpretation of every worked comparison.

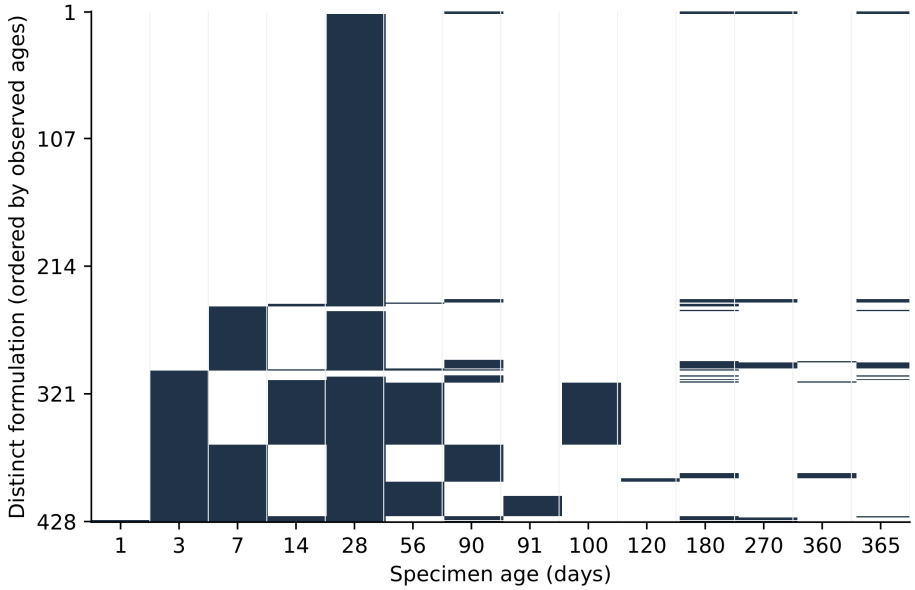


Figure 1: Observed ages across the 428 numerical formulations. Dark cells mark available formulation–age observations; blank cells mark absent observations. Rows are ordered by their age-availability pattern, not by strength. Distances between columns do not represent equal time intervals.

2.2. Defining an admissible reversal

For formulation i, let $s_i(t)$ denote its descriptive strength at age t. The signed contrast for a pair ordered by identifier is

$$d_{ij}(t) = s_i(t) - s_j(t), \qquad i < j. \tag{1}$$

A comparison is admissible when both ages are observed and its smaller absolute contrast reaches a chosen margin δ:

$$A_{ij} = \mathbf{1}\left\{\min\left(|d_{ij}(t_0)|, |d_{ij}(t_1)|\right) \geq \delta\right\}. \tag{2}$$

Eq. (1) fixes the direction of subtraction, while Eq. (2) excludes comparisons whose displayed separation is small at either age. A reversed ordering satisfies

$$R_{ij} = \mathbf{1}\{d_{ij}(t_0)d_{ij}(t_1) < 0\}. \tag{3}$$

The ordering change in Eq. (3) concerns two observations. It does not estimate a crossing time between them. The contrast magnitude used to order cases is

$$q_{ij} = \min\left(|d_{ij}(t_0)|, |d_{ij}(t_1)|\right). \tag{4}$$

The principal margin is 3 MPa, accompanied by checks at one and five megapascals. This is a display and selection choice, not an estimate of measurement uncertainty or a prescribed acceptance criterion for concrete. Additional comparisons use three to twenty-eight, twenty-eight to ninety, and seven to ninety days. Eq. (4) also permits examination of the full range of contrast magnitudes instead of treating the chosen margin as a natural boundary in material behavior.

2.3. Selecting cases under formulation exclusion

Reversal-prioritized selection orders admissible pairs first by reversal status and then by descending contrast magnitude. Identifier order resolves ties. A second rule orders pairs by the largest absolute early-age contrast, retaining

the same tie rule. Random ordering shuffles all admissible pairs. Each rule scans its ordering, accepts a pair only when neither formulation has already been used, and stops after ten accepted cases. The ten-case length is an instructional allocation, not a conclusion that ten comparisons are sufficient for mastery. Requiring distinct formulations prevents a single conspicuous recipe from supplying repeated demonstrations.

Random ordering is evaluated a thousand times with a NumPy generator initialized at seed 20261026. The age intervals are processed in the order seven to twenty-eight, three to twenty-eight, twenty-eight to ninety, and seven to ninety days, each at margins of one, three and five megapascals. The generator advances through that fixed order. Counts from these sessions describe alternative arrangements of the same observations. Their central percentile range concerns scheduling variability, not uncertainty about a population of concretes. All three rules inspect an admissible pool already defined using both ages, so the comparison concerns case ordering rather than prediction from early evidence alone.

Shared-formulation dependence is examined by counting how many admissible reversal partners each formulation possesses. Figure 2 makes this dependence visible without depicting pair counts as independent experimental samples. A formulation with many partners contributes to many candidate comparisons, yet it may still enter only one accepted case. The exclusion rule therefore changes the practical meaning of a large candidate count. It also creates the possibility that a greedy ordering consumes a formulation that would have supported several otherwise compatible choices.

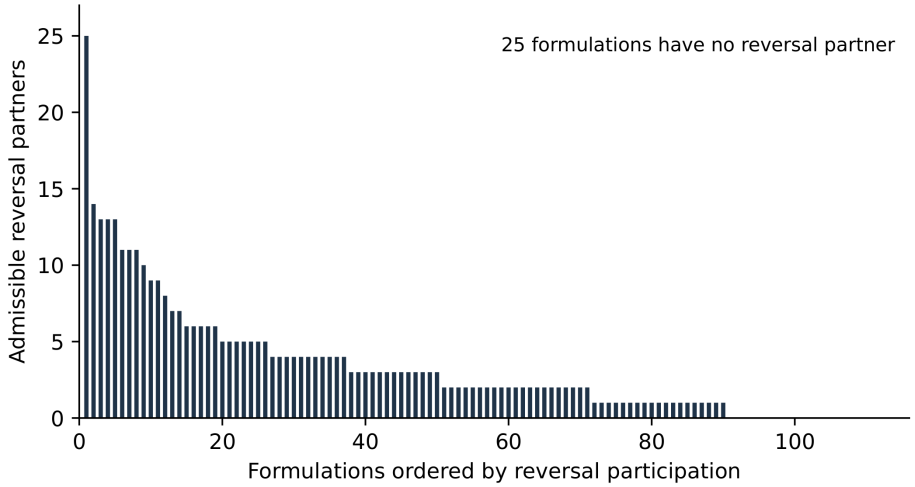


Figure 2: Reversal partners per formulation in the seven-to-twenty-eight-day comparison at a 3 MPa margin. Each vertical mark represents one of the 115 eligible formulations, ordered by partner count. Repeated participation explains why comparison counts cannot be interpreted as independent sample sizes.

2.4. Exact attainable selection

An integer calculation determines how many reversal pairs can coexist without sharing a formulation. Each admissible reversal is an edge in a comparison graph, with one binary decision x_e indicating whether that pair is selected. For each formulation v, the sum of incident selected pairs cannot exceed one:

$$C = \max_{x_e \in \{0,1\}} \sum_{e \in E_R} x_e, \qquad \sum_{e \ni v} x_e \le 1 \quad \text{for every } v. \tag{5}$$

Here E_R contains only the admissible reversed pairs. Eq. (5) gives their maximum cardinality under formulation exclusion. The attainable reversal count for a session of length B is

$$C_B = \min(B, C). \tag{6}$$

The integer solver terminated successfully with zero relative optimality tolerance for every age-and-margin combination. Each selected optimum was also checked for completion to ten cases using other admissible pairs when fewer than ten reversals were available. The corresponding selected pairs were checked against the incidence constraints.

Eq. (6) concerns the number of reversals that can be chosen; it does not identify the best educational ordering among equally large selections. It supplies an exact comparison for detecting avoidable losses in the greedy rule.

2.5. Offline interaction and numerical checks

The instructional resource displays both compositions and the early-age strengths with units. Learners choose which formulation they expect to be stronger at the later age, or choose an undetermined ordering, and enter a reason. The later strengths remain concealed until that response is committed. The revealed view retains the earlier information, presents the later comparison, and asks for a statement explicitly qualified by age. A second text response records that statement. The resource stores responses in memory during the session and permits local export; closing the activity clears unexported responses. Inspection of its code can reveal the later values, so this delivery mechanism is intended for learning rather than secure examination.

A demonstration sequence contains the principal rule's ten reversed comparisons. A separate practice sequence interleaves five reversals and five retained orderings without repeating formulations. The second sequence reduces the likelihood that reversal becomes a predictable answer cue. Introductory guidance explains the units, the meaning of age, and the difference between an expected answer and a conclusion established by the visible evidence. Each text field requires at least ten characters before progression, which verifies entry rather than scientific quality. Feedback identifies the observed ordering and the limitation on an ingredient-specific explanation; it does not assign an automated grade to the reasoning.

This restrained presentation follows evidence that reducing irrelevant information and coordinating displayed material can support the management of cognitive demands [27]. Research on immersive learning also indicates that a convincing visual environment and an effective learning process are different design concerns [28]. Accordingly, the activity uses numerical tables instead of depicting an unobserved laboratory procedure. Graphics use distinguishable tones, restrained color and explicit units, consistent with evidence about misleading color choices in scientific communication [29].

The principal pair counts were checked through an independent matrix-of-differences calculation. The selected principal cases were verified to contain twenty different identifiers and to meet the required contrast at both ages. Eighty displayed formulation–age entries across the two sequences were compared with their corresponding numerical records, with agreement in every check. Functional evaluation covered both ten-case sequences, prediction gating, reflection gating and export using a controlled document-object stand-in. Browser rendering and cross-device operation were not verified. Additional calculations replace the within-cell median by the mean, exclude cells containing two distinct strengths, and regroup recipes after rounding ingredient quantities to one decimal place.

3. RESULTS AND DISCUSSION

3.1. Available ages and the scale of the comparison task

The eligible principal formulations produce 6,555 unordered pairs before applying the contrast margin. Of these, 5,141 remain admissible at 3 MPa, and 185 reverse their measured ordering. The reversal share is therefore 3.60 percent within the admitted pool. This percentage refers to a set of comparisons constructed from the measurements, not to the proportion of concrete mixtures expected to reverse in another laboratory. Its denominator changes when the age interval or margin changes. The teaching opportunity is nevertheless concrete: the collection contains enough clearly separated counterexamples to an age-independent ranking to populate a short session without inventing a hypothetical material response.

The unequal age coverage in Figure 1 helps explain why comparison counts differ. Twenty-eight-day observations are present for 417 formulations, whereas seven-day observations are available for 119 and ninety-day observations for fifty-four. The required intersections are smaller still. The principal comparison retains 115 formulations, the three-to-twenty-eight-day interval retains 122, and the twenty-eight-to-ninety-day interval retains fifty-one. A long interval does not automatically offer more usable cases, because the corresponding ages must coexist for each recipe.

Observation coverage therefore determines which instructional questions the measurements can support before any selection policy is applied.

These distinctions matter for classroom interpretation. A table containing many strengths at a conventional reference age can appear extensive while providing little support for a question about change over time. Displaying the availability pattern makes the requirement for matched ages explicit. It also prevents a convenient but unjustified substitution of ninety-one days for ninety days. Such substitution could increase the number of cases, but it would change the comparison and would need its own rationale. Keeping the ages separate allows the instructor to state exactly which observations support a ranking rather than implying a continuous strength history from sparse tests.

Table 2 places the comparison supply alongside the number of compatible reversals. The distinction between a pair count and a compatible selection is essential. An instructor who needs ten different comparisons cannot simply choose the ten largest reversals if they share formulations. Conversely, a relatively small reversal pool may still support a complete session if its pairs involve different recipes. The calculations therefore examine both the amount of reversal evidence and its combinatorial arrangement. Neither calculation estimates the amount of teaching time required or the number of examples a particular student needs.

Table 2: Comparison supply and selection under formulation exclusion.

Ages (days)	δ (MPa)	Recipes	Admissible	Reversed	Exact	Greedy	Random
7–28	1	115	5,950	456	51	10	0.735
7–28	3	115	5,141	185	30	10	0.314
7–28	5	115	4,483	80	16	10	0.178
3–28	1	122	6,618	802	61	10	1.206
3–28	3	122	5,410	352	43	10	0.657
3–28	5	122	4,499	131	24	10	0.289
28–90	1	51	1,139	75	16	10	0.665
28–90	3	51	935	14	6	6	0.155
28–90	5	51	792	2	1	1	0.027
7–90	1	43	801	159	19	10	1.997
7–90	3	43	608	84	13	10	1.494
7–90	5	43	465	48	9	6	1.091

3.2. Dependence through repeated participation

Ninety principal-window formulations have at least one admissible reversal partner, while twenty-five have none. Among the participating formulations, the median number of partners is three and the largest number is twenty-five. The ordered marks in Figure 2 reveal the uneven participation. Some recipes can generate many conspicuous comparisons, whereas others contribute only a single opportunity. A count of reversed pairs therefore overstates the number of distinct material experiences if the activity repeatedly returns to the same influential recipe. Formulation exclusion addresses that instructional concentration directly, without asserting that the remaining recipes are independent experimental batches.

The distinction also changes how numerical evidence should be summarized. Treating every pair as an independent observation would allow a single strength value to influence many entries while concealing the shared origin of those entries. More pairs can improve the inventory of possible teaching contrasts, but they do not create more physical tests. The analysis therefore does not attach an independent-sample standard error to the reversal proportion. It also avoids a significance test comparing the targeted and random schedules, since the targeted rule is constructed to prefer known reversals. The meaningful question is what each rule delivers under the same restrictions.

Maximum compatible selection gives a further interpretation of the principal pool. Thirty reversal pairs can be selected without sharing a formulation, using sixty recipes. The ten-case demonstration therefore uses only part of the available compatible supply. This offers flexibility for constructing alternative teaching sessions, but it does not establish that thirty cases should be assigned. A longer activity may become repetitive or place excessive

demands on attention. The count establishes availability; selecting the amount and order of instruction remains an educational decision that must consider prior knowledge, the intended explanation and the time available for discussion.

The dependence analysis also suggests a useful way to read a student's repeated success. If several questions reuse the same recipe, recognition of its identifier or remembered strength can replace comparison of conditions. A session using different recipes reduces that particular shortcut. It cannot remove every possible cue, because different formulations may still have similar ingredient patterns or originate from related batches. In any later test of transfer, the separation between practice and assessment should therefore operate at the formulation level and, where physical provenance permits, at the batch level. That requirement follows from the measurement structure rather than from a preference for a more elaborate interface.

3.3. Selection-order losses and limits in the available evidence

Figure 3 compares greedy reversal selection with the exact attainable count for a ten-case session. In the principal interval, every examined margin supports ten compatible reversals. The uncapped exact counts fall from fifty-one at the smallest margin to thirty at the principal margin and sixteen at the strictest. The greedy rule reaches the session limit throughout, so exact optimization offers no increase in the number delivered for these settings. It still provides a useful check: the apparent success of greedy selection is confirmed against an attainable count rather than assumed from the size of the pair pool.

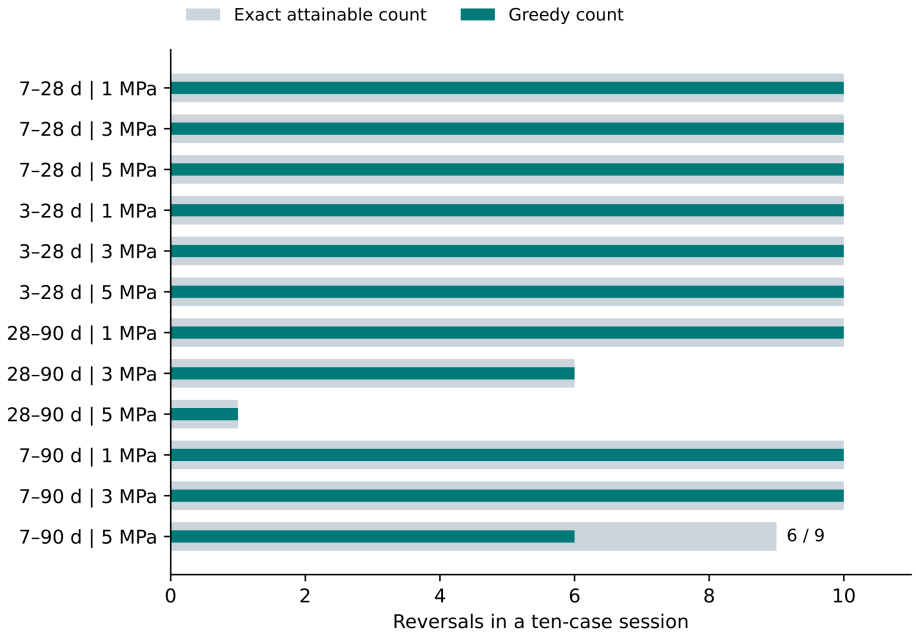


Figure 3: Attainable and greedily selected reversals for a ten-case session.

The twenty-eight-to-ninety-day interval illustrates a different limitation. At the principal margin, fourteen reversed pairs are available, yet at most six can coexist under formulation exclusion. The greedy rule selects six, reaching the exact limit. Increasing the margin to five megapascals leaves two reversed pairs that can support only one compatible choice. A different ordering cannot create a ten-reversal session from those observations. The constraint comes from which recipes share the available comparisons, not simply from a poor choice of ordering. An instructor seeking additional examples would have to relax the margin, change the ages or obtain further measurements.

For seven to ninety days at the strictest margin, the greedy rule selects six reversals even though nine are compatible. This difference identifies an avoidable selection loss. Favoring the largest contrast can consume formulations that would support several compatible alternatives. Exact cardinality selection avoids that loss by considering the whole set of exclusion constraints. The result does not show that its nine cases are educationally superior to the six

chosen greedily: some may have smaller contrasts, less useful ingredient variation or explanations that require more background knowledge. It establishes that the greedy count is not the available limit.

The uncapped capacity also shows why a fixed session budget can hide differences between selection settings. At the principal margin, the exact solution uses sixty of the ninety formulations that participate in at least one reversal. At the strictest principal-window margin, the compatible supply remains larger than the classroom allocation even though it is much smaller than at the lowest margin. All three settings therefore appear equally successful if only the ten-case completion count is inspected. An instructor preparing several different sessions would encounter their differences much sooner. The uncapped count is useful for that preparation question, whereas the capped count directly describes one session. Neither should be substituted for the other when discussing available teaching variety.

The practical role of the integer calculation is therefore diagnostic. It distinguishes an authoring choice from an evidential restriction. When the greedy and exact counts agree, efforts to increase the number of reversals through reordering alone cannot succeed. When they differ, an instructor can inspect the alternative compatible selection and judge whether the additional cases serve the intended lesson. Counting remains subordinate to the instructional purpose. A selection with more examples is useful only when learners can examine the associated conditions and formulate a defensible comparison, rather than being encouraged to accumulate correct guesses.

3.4. What random ordering exposes

The principal random sessions contain zero, one, two or three reversals. Their respective frequencies are 732, 227, thirty-six and five, as shown in Figure 4. The mean count is 0.314, and the central empirical percentile range extends from zero to two. The chance of encountering at least one reversal in these particular randomized sessions is 26.8 percent. This is a property of the implemented shuffling and exclusion procedure applied to the fixed comparison pool. It should not be interpreted as a student's probability of recognizing a reversal or as an estimate for another measurement collection.

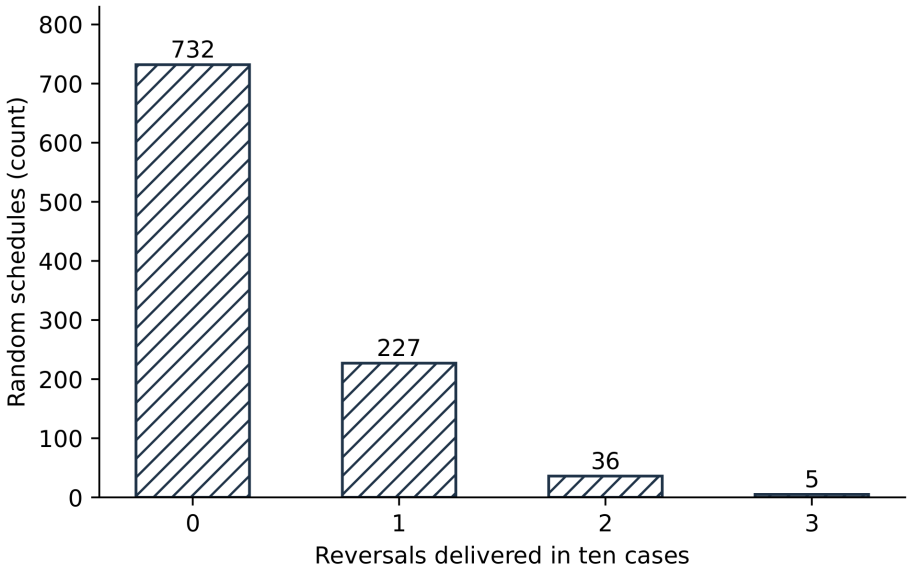


Figure 4: Discrete reversal counts across the principal random sessions. Bar labels give the number of ten-case sessions in each category. The distribution describes shuffled delivery with formulation exclusion; it contains no student-performance observations.

The results explain why an apparently adequate collection of material observations may fail to expose the intended contrast during a brief activity. Reversals occupy only a small part of the admissible principal pool, and accepting a pair removes its formulations from subsequent choices. Multiplying the pool's reversal share by ten would give approximately 0.360 reversals, but that calculation does not reproduce the actual selection mechanism. It ignores the change in available pairs after each acceptance. The empirical distribution is preferable for describing this delivery

rule because it respects the implemented sequence of restrictions without assigning an independence assumption to the cases.

Selecting by the largest early-strength separation exposes no reversal in the principal session. This does not mean that large early differences are poor teaching examples in general. They may be useful for learning units, comparing magnitudes or discussing a clearly separated response at one age. They are poorly aligned with the narrower purpose of demonstrating a later change in ordering. The comparison thus illustrates a difference between the intended lesson and a convenient numerical sorting rule. The authoring objective should be stated before selecting examples so that visual clarity at the initial age is not mistaken for relevance to the later-age question.

The advantage of targeted selection is expected because the later strengths are known during preparation. Presenting that advantage as an unknown-outcome prediction result would be invalid. Research on leakage in machine-learning-based science explains why access to information unavailable at prediction time can lead to exaggerated conclusions [30]. Here that access is legitimate for choosing instructional cases and must remain explicit. Research evaluating material-property prediction methods addresses a separate requirement: performance on an appropriately separated prediction task [31]. The present counts answer an authoring question and do not compete with predictive error estimates.

3.5. Contrast magnitude and sensitivity to numerical handling

The smaller endpoint contrast provides a useful description of how strongly a reversal is separated at both ages. Figure 5 displays its distribution for all reversed pairs before imposing a positive margin. The medians are approximately 1.85 MPa for three to twenty-eight days, 1.50 MPa for seven to twenty-eight days, 1.21 MPa for twenty-eight to ninety days, and 2.34 MPa for seven to ninety days. These values show that many sign changes have a small contrast at one endpoint. A sign reversal alone therefore does not guarantee a visually clear comparison for classroom discussion.

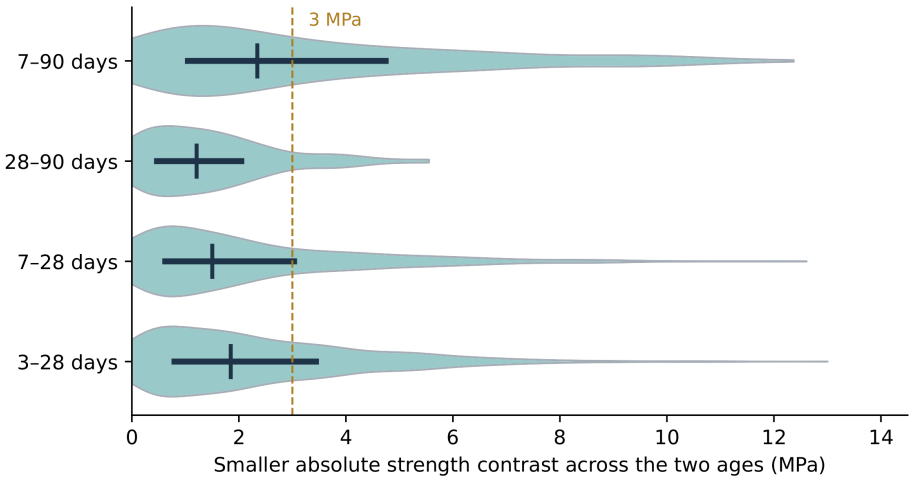


Figure 5: Distribution of the smaller absolute endpoint contrast among reversed comparisons. The horizontal extent summarizes the observed contrast magnitudes for each age interval; thick marks indicate the interquartile range and central marks the median. The dashed reference marks 3 MPa. Smoothed outlines aid description and do not represent inferential probability intervals.

Increasing the principal-window margin from one to five megapascals reduces the reversal inventory from 456 to eighty. The analogous decrease for three to twenty-eight days is from 802 to 131. By contrast, the twenty-eight-to-ninety-day interval has a maximum smaller-endpoint contrast of only about 5.55 MPa, leaving very little supply at the strictest examined margin. The distribution helps explain the exact-selection restrictions without asserting that small contrasts are measurement errors. Their uncertainty cannot be quantified from the available metadata. The margin controls what enters an instructional comparison; it cannot establish which observed difference is physically reliable.

Uncertainty-aware strength estimation provides an important related perspective. Conformal approaches have been developed to provide intervals rather than only point predictions for concrete strength [32]. Those intervals answer a different question from the display margin used here. The present analysis neither fits a strength predictor nor calibrates a distribution of test error. Labeling a three-megapascal difference as statistically secure would consequently be unjustified. The same caution applies when a large reversal is visually compelling: an effect size in the displayed numbers does not repair absent information about curing, specimen geometry or independent replication.

Changing the within-cell median to the mean leaves the principal counts unchanged because each multi-valued cell contains only two strengths. This equality is arithmetic rather than strong evidence of robustness. A more discriminating check excludes those cells, retaining 112 eligible formulations, 4,837 admissible comparisons and 163 reversals. Regrouping recipes after rounding ingredient quantities to one decimal place gives 427 combinations overall and 114 eligible principal formulations, with 5,040 admissible comparisons and 179 reversals. Both checks retain enough compatible contrasts for the principal demonstration. They show how numerical identity and aggregation affect case construction, while leaving the missing physical conditions unresolved.

3.6. Movement in formulation rank

A change in pairwise ordering can also be examined through the position of each formulation among all eligible recipes. Using average ranks for ties and placing the strongest formulation first, fifty-two formulations improve their position between the principal ages, fifty-seven move downward and six retain the same position. The median absolute change is seven rank positions, while the largest absolute change is fifty-three. Figure 6 shows the distribution of these movements. This formulation-level view complements the pairwise analysis by showing that reordered comparisons arise from a broader rearrangement rather than from a single exceptional pair.

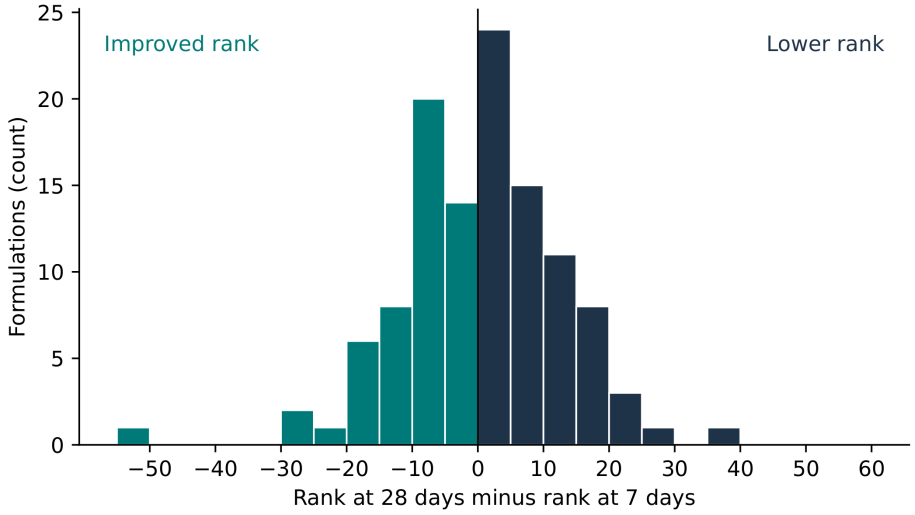


Figure 6: Rank displacement across the 115 principal formulations. Negative changes indicate an improved position at twenty-eight days and positive changes a lower position; ties receive average ranks. The histogram concerns relative ordering, not absolute strength loss.

A downward rank movement does not mean that a formulation loses absolute strength. Rank depends on the other recipes included in the comparison, so a mixture may become stronger while others strengthen more. The distinction is useful educationally because students may conflate relative standing with material deterioration. The activity can make both quantities available: the measured strength at each age and the ordering among eligible recipes. A defensible interpretation should identify which quantity has changed. Neither a favorable rank nor a large absolute strength by itself establishes suitability for a particular design requirement or service environment.

The rank distribution is conditional on the eligible formulation set. Removing recipes, adding new measurements or changing the age window can alter rank movement even when the retained strength values remain fixed. It is therefore inappropriate to treat the observed rank shifts as universal attributes of the formulation identifiers. The

labels serve only to connect composition and measured responses during the activity. This qualification reinforces the central lesson that a comparison is attached to its conditions. The relevant conditions include the measured age and, for a relative position, the set against which the formulation is being compared.

3.7. A material comparison with several simultaneous changes

The first selected comparison pairs M357 with M399. At seven days, the descriptive strengths are 39.40 and 52.01 MPa, respectively. At twenty-eight days, they are 71.30 and 57.03 MPa. The signed contrast therefore changes from approximately negative 12.61 MPa to positive 14.27 MPa. Both mixtures gain absolute strength over the interval, but their gains differ sufficiently to reverse the ordering. The supported statement identifies the stronger mixture separately at each measured age. It does not require the learner to estimate an intermediate crossing time or assume a particular continuous growth curve.

Composition differences complicate a causal explanation. Figure 7 shows that M357 contains 83.40 kilograms per cubic metre less cement, 165.00 more slag and 79.00 less fly ash than M399. Water and both aggregate quantities also differ, while superplasticizer differs only slightly. Table 3 retains the absolute quantities so that the signed differences can be checked in their physical context. A learner who attributes the reversal solely to the slag quantity overlooks the other changes and the absent curing and specimen information. The comparison establishes a change in measured ordering, not an isolated constituent effect.

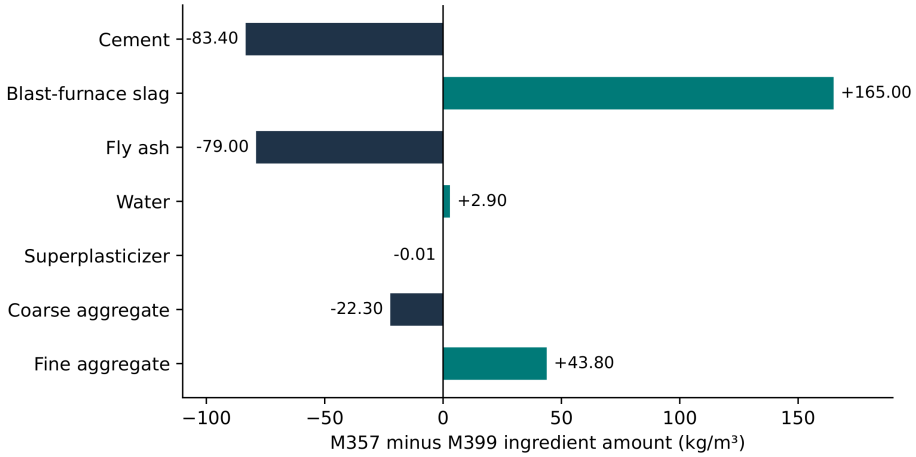


Figure 7: Ingredient changes within the worked comparison, expressed as M357 minus M399. The diverging bars expose several simultaneous changes, preventing the strength reversal from being visually assigned to a single constituent. Values retain the units of material quantity per cubic metre.

Table 3: Absolute quantities for the worked comparison.

Quantity	M357	M399
Cement	362.60	446.00
Blast-furnace slag	189.00	24.00
Fly ash	0.00	79.00
Water	164.90	162.00
Superplasticizer	11.60	11.61
Coarse aggregate	944.70	967.00
Fine aggregate	755.80	712.00
Total binder	551.60	549.00
Water-to-binder mass ratio	0.299	0.295
Strength at seven days	39.40	52.01
Strength at twenty-eight days	71.30	57.03

Summing cement, slag and fly ash gives total binder quantities of 551.60 and 549.00 kilograms per cubic metre for the two mixtures. Dividing water by these totals gives mass ratios of approximately 0.299 and 0.295. These ratios

are calculated from ingredient masses and introduce no additional strength observations or assumptions about reaction completion. These derived quantities show that the comparison involves nearly equal total binder and similar water-to-binder ratios despite the large change in constituent balance. A learner who compares only the cement amounts would miss that similarity, while a learner who compares only total binder would miss the different distribution among cement, slag and fly ash. The example therefore supports examining several descriptions of composition before interpreting the strength ordering.

The derived ratios do not establish equivalent reaction progress, porosity or effective binder activity. Counting a kilogram of fly ash or slag within total binder does not imply that it contributes the same reaction products at the same age as a kilogram of cement. Without reactivity measurements, curing information and specimen identification, the numerical similarity cannot distinguish those possibilities. It is useful as an additional observation to explain, rather than as a correction that makes the comparison controlled. This prevents a convenient composition ratio from replacing the missing material evidence.

Research on supplementary cementitious reactivity illustrates what additional evidence would be needed for a mechanistic argument. Calorimetric and thermogravimetric approaches can distinguish reactions that ingredient quantities alone cannot resolve [33]. A rapid reactivity test for calcined clays likewise links a defined material and controlled test procedure to an assessed responsen [34]. Those techniques are relevant examples of experimental discrimination; they were not performed on the two formulations examined here. Their role in the discussion is to identify the difference between a plausible material explanation and evidence capable of testing that explanation.

The worked comparison also clarifies why selecting a dramatic case is not sufficient instruction. The numerical surprise should lead to a more carefully qualified claim, rather than an unsupported story about one ingredient. Asking what changed, what remained visible, and which missing condition would help distinguish explanations gives the teacher a specific basis for discussion. The activity preserves the earlier response so that this change in reasoning can be examined. It does not assume that a changed prediction demonstrates understanding, since a learner could simply repeat the newly displayed ordering without acknowledging the conditions that limit it.

3.8. Instructional interpretation and the next empirical test

Feedback in the activity identifies the observed ordering and asks the learner to qualify the conclusion by age. Educational-feedback research indicates that the effect of feedback depends on its content and instructional role rather than its presence alone [35]. Accordingly, displaying the later number is only part of the interaction. The written response should connect that observation to the scope of the preceding claim. A response that merely names the later winner is incomplete for this purpose, even when it predicts correctly. This criterion gives an instructor a more specific target than general participation or satisfaction.

The mixed practice sequence creates a further distinction between demonstrating a principle and testing whether it has been understood. A demonstration can deliberately select several reversals to establish that a universal claim fails. Practice should include retained orderings so that the learner must examine the numbers rather than infer the intended answer from the activity type. The undetermined option serves a related purpose before disclosure: it permits an explicit acknowledgment that the early comparison does not settle the later one. After disclosure, that answer should be replaced by a statement about the observed ages, while uncertainty about unmeasured conditions remains. The two responses concern different evidential states. Treating them as equivalent would obscure the very change that the delivery mechanism is designed to expose.

The design also separates immediate comparison from durable learning. Evidence concerning retrieval practice and spacing supports assessing whether knowledge can be used after the learning episode rather than only while the relevant values remain visible [36, 37]. A later assessment could therefore use previously unseen formulations and ask for an explanation of which age-conditioned claims are warranted before and after disclosure. It should not reward memorization of the two numerical strengths in the demonstration. No retention interval or transfer outcome was measured here, so these possibilities define an empirical evaluation rather than extend the present calculations into educational effects.

Learning analytics research in virtual laboratories emphasizes the need to relate interaction records to meaningful educational questions [38]. Completion, prediction changes and exported text would supply process information, but they would not establish why a learner changed an answer. Human assessment of explanations would be required to distinguish recognition of age dependence from a guess or a copied phrase. Prior knowledge and the balance of stable and reversed cases would also need attention. An exclusively reversal-based sequence could train an answer pattern that succeeds within the activity and fails when the material evidence is less predictable.

The technical checks place additional limits on implementation claims. Correctly gating a response in a controlled software check establishes that the intended condition can be evaluated by the program. It does not establish that a learner notices the prompt, understands the units, can comfortably enter a response on a particular device or receives an accessible announcement after disclosure. Those questions require direct observation and accessibility testing. The simple screen arrangement reduces the amount of material displayed, but its convenience has not been measured. Retaining these distinctions prevents operational success from being used as a substitute for evidence about learners or the conditions under which they use the resource.

A direct educational comparison should therefore hold the compositions, ages, case count and explanations constant while varying when the later values become visible. Immediate disclosure and prediction-before-disclosure could then be compared without confounding the amount of material evidence. Explanations should be assessed by raters unaware of the delivery condition, with explicit attention to units, the age attached to each ordering and the refusal to assign a single-ingredient cause without supporting evidence. The present work supplies compatible cases for that comparison. It does not establish which delivery condition would produce better understanding or how the answer would vary across students.

4. CONCLUSIONS

Offline materials instruction can deliver age-conditioned strength comparisons while requiring learners to distinguish an expectation from an observation-supported claim. The laboratory measurements support a substantial inventory of reversed orderings, but shared formulations constrain how many can appear together without repetition. Exact selection establishes when a short sequence is limited by the available comparisons and when a greedy ordering leaves compatible examples unused. This distinction turns a general request for contrasting material evidence into a specific preparation decision.

The graphical analyses show complementary features of that decision: unequal observation coverage, repeated participation by influential formulations, the attainable number of compatible cases, the low reversal exposure under random ordering, the concentration of weak endpoint contrasts, changes in relative rank and simultaneous ingredient differences. Together they support a narrower and more defensible use of numerical evidence in teaching. They do not establish a controlled material mechanism or an educational treatment effect. The appropriate next investigation is an assessment of age-qualified reasoning using the same material comparisons under controlled disclosure timing, with unfamiliar formulations reserved for transfer.

Author Contributions

All authors contributed to the conception and design of the study, analysis and interpretation of the results, and manuscript preparation. All authors reviewed and approved the final manuscript.

Conflict of Interests

The authors declare that they have no conflicts of interest.

Funding Information

This research received no external funding.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. M.K. Swenty, B.Z. Dymond, J.C. Carroll, Effective Teaching Methods in Concrete Education, *ACI Special Publication* **359**, 1–17 (2023), https://doi.org/10.14359/51740287.
2. K.L. Scrivener, P. Juilland, P.J.M. Monteiro, Advances in understanding hydration of Portland cement, *Cement and Concrete Research* **78**, 38–56 (2015), https://doi.org/10.1016/j.cemconres.2015.05.025.
3. M.C.G. Juenger, R. Snellings, S.A. Bernal, Supplementary cementitious materials: New sources, characterization, and performance insights, *Cement and Concrete Research* **122**, 257–273 (2019), https://doi.org/10.1016/j.cemconres.2019.05.008.
4. G. Habert, S.A. Miller, V.M. John, J.L. Provis, A. Favier, A. Horvath, *et al.*, Environmental impacts and decarbonization strategies in the cement and concrete industries, *Nature Reviews Earth & Environment* **1**, 559–573 (2020), https://doi.org/10.1038/s43017-020-0093-3.
5. K.L. Scrivener, V.M. John, E.M. Gartner, Eco-efficient cements: Potential economically viable solutions for a low-CO2 cement-based materials industry, *Cement and Concrete Research* **114**, 2–26 (2018), https://doi.org/10.1016/j.cemconres.2018.03.015.
6. D. Vergara, M.P. Rubio, M. Lorenzo, New Approach for the Teaching of Concrete Compression Tests in Large Groups of Engineering Students, *Journal of Professional Issues in Engineering Education and Practice* **143**, 05016009 (2017), https://doi.org/10.1061/(asce)ei.1943-5541.0000311.
7. V. Rodríguez-Hernández, M.C. Espino-Gudiño, J. Gudiño-Bazaldua, J.L. González-Pérez, V.M. Castaño, Adapting Learning Objects to E-Learning and B-Learning in Materials Science Curricula, *Journal of Materials Education* **34**, 29–44 (2012), https://icme.unt.edu/files/34_1_2.pdf.
8. S.M. Reeves, K.J. Crippen, Virtual Laboratories in Undergraduate Science and Engineering Courses: a Systematic Review, 2009–2019, *Journal of Science Education and Technology* **30**, 16–30 (2021), https://doi.org/10.1007/s10956-020-09866-0.
9. P. Chan, T. Van Gerven, J.L. Dubois, K. Bernaerts, Virtual chemical laboratories: A systematic literature review of research, technologies and instructional design, *Computers and Education Open* **2**, 100053 (2021), https://doi.org/10.1016/j.caeo.2021.100053.
10. C.C. Chase, L. Malkiewich, A. S. Kumar, Learning to notice science concepts in engineering activities and transfer situations, *Science Education* **103**, 440–471 (2019), https://doi.org/10.1002/sce.21496.
11. C.R. Bego, R.J. Chastain, M.S. DeCaro, Designing novel activities before instruction: Use of contrasting cases and a rich dataset, *British Journal of Educational Psychology* **93**, 299–317 (2023), https://doi.org/10.1111/bjep.12555.
12. K. Loibl, I. Roll, N. Rummel, Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning, *Educational Psychology Review* **29**, 693–715 (2017), https://doi.org/10.1007/s10648-016-9379-x.
13. T. Sinha, M. Kapur, When Problem Solving Followed by Instruction Works: Evidence for Productive Failure, *Review of Educational Research* **91**, 761–798 (2021), https://doi.org/10.3102/00346543211019105.
14. K. Loibl, M. Tillema, N. Rummel, T. van Gog, The effect of contrasting cases during problem solving prior to and after instruction, *Instructional Science* **48**, 115–136 (2020), https://doi.org/10.1007/s11251-020-09504-7.
15. G. Brod, Predicting as a learning strategy, *Psychonomic Bulletin & Review* **28**, 1839–1847 (2021), https://doi.org/10.3758/s13423-021-01904-1.
16. G. Brod, A. Greve, D. Jolles, M. Theobald, E.M. Galeano-Keiner, Explicitly predicting outcomes enhances learning of expectancy-violating information, *Psychonomic Bulletin & Review* **29**, 2192–2201 (2022), https://doi.org/10.3758/s13423-022-02124-x.
17. L. Fiorella, Making Sense of Generative Learning, *Educational Psychology Review* **35**, 50 (2023), https://doi.org/10.1007/s10648-023-09769-7.
18. S.C. Pan, F. Sana, Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice, *Journal of Experimental Psychology: Applied* **27**, 237–257 (2021), https://doi.org/10.1

037/xap0000345.
19. L. Deslauriers, L.S. McCarty, K. Miller, K. Callaghan, G. Kestin, Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom, *Proceedings of the National Academy of Sciences* **116**, 19251–19257 (2019), https://doi.org/10.1073/pnas.1821936116.
20. E.J. Theobald, M.J. Hill, E. Tran, S. Agrawal, E.N. Arroyo, S. Behling, *et al.*, Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math, *Proceedings of the National Academy of Sciences* **117**, 6476–6483 (2020), https://doi.org/10.1073/pnas.1916903117.
21. B.A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods, *Cement and Concrete Research* **115**, 379–388 (2019), https://doi.org/10.1016/j.cemconres.2018.09.006.
22. P.G. Asteris, A.D. Skentou, A. Bardhan, P. Samui, K. Pilakoutas, Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models, *Cement and Concrete Research* **145**, 106449 (2021), https://doi.org/10.1016/j.cemconres.2021.106449.
23. Y. Zhang, W. Ren, Y. Chen, Y. Mi, J. Lei, L. Sun, Predicting the compressive strength of high-performance concrete using an interpretable machine learning model, *Scientific Reports* **14**, 28346 (2024), https://doi.org/10.1038/s41598-024-79502-z.
24. S.I. Hassan, S.A. Syed, S.W. Ali, H. Zahid, S. Tariq, M. Mohd Su ud, *et al.*, Systematic literature review on the application of machine learning for the prediction of properties of different types of concrete, *PeerJ Computer Science* **10**, e1853 (2024), https://doi.org/10.7717/peerj-cs.1853.
25. I.C. Yeh, *Concrete Compressive Strength*, UCI Machine Learning Repository, accession 165 (1998), https://doi.org/10.24432/C5PK67.
26. I.C. Yeh, Modeling of strength of high-performance concrete using artificial neural networks, *Cement and Concrete Research* **28**, 1797–1808 (1998), https://doi.org/10.1016/s0008-8846(98)00165-3.
27. J.C. Castro-Alonso, B.B. de Koning, L. Fiorella, F. Paas, Five Strategies for Optimizing Instructional Materials: Instructor- and Learner-Managed Cognitive Load, *Educational Psychology Review* **33**, 1379–1407 (2021), https://doi.org/10.1007/s10648-021-09606-9.
28. G. Makransky, G.B. Petersen, The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality, *Educational Psychology Review* **33**, 937–958 (2021), https://doi.org/10.1007/s10648-020-09586-2.
29. F. Crameri, G.E. Shephard, P.J. Heron, The misuse of colour in science communication, *Nature Communications* **11**, 5444 (2020), https://doi.org/10.1038/s41467-020-19160-7.
30. S. Kapoor, A. Narayanan, Leakage and the reproducibility crisis in machine-learning-based science, *Patterns* **4**, 100804 (2023), https://doi.org/10.1016/j.patter.2023.100804.
31. A. Dunn, Q. Wang, A. Ganose, D. Dopp, A. Jain, Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm, *npj Computational Materials* **6**, 138 (2020), https://doi.org/10.1038/s41524-020-00406-3.
32. P. Tamuly, V. Nava, Machine learning based conformal predictors for uncertainty-aware compressive strength estimation of concrete, *Construction and Building Materials* **487**, 141844 (2025), https://doi.org/10.1016/j.conbuildmat.2025.141844.
33. P. Suraneni, A. Hajibabaee, S. Ramanathan, Y. Wang, J. Weiss, New insights from reactivity testing of supplementary cementitious materials, *Cement and Concrete Composites* **103**, 331–338 (2019), https://doi.org/10.1016/j.cemconcomp.2019.05.017.
34. F. Avet, R. Snellings, A. Alujas Diaz, M. Ben Haha, K. Scrivener, Development of a new rapid, relevant and reliable (R3) test method to evaluate the pozzolanic reactivity of calcined kaolinitic clays, *Cement and Concrete Research* **85**, 1–11 (2016), https://doi.org/10.1016/j.cemconres.2016.02.015.
35. B. Wisniewski, K. Zierer, J. Hattie, The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research, *Frontiers in Psychology* **10**, 3087 (2020), https://doi.org/10.3389/fpsyg.2019.03087.
36. S.K. Carpenter, S.C. Pan, A.C. Butler, The science of effective learning with spacing and retrieval practice, *Nature Reviews Psychology* **1**, 496–511 (2022), https://doi.org/10.1038/s44159-022-00089-1.

37. P.K. Agarwal, L.D. Nunes, J.R. Blunt, Retrieval Practice Consistently Benefits Student Learning: a Systematic Review of Applied Research in Schools and Classrooms, *Educational Psychology Review* **33**, 1409–1453 (2021), https://doi.org/10.1007/s10648-021-09595-9.
38. R. Elmoazen, M. Saqr, M. Khalil, B. Wasson, Learning analytics in virtual laboratories: a systematic literature review of empirical research, *Smart Learning Environments* **10**, 23 (2023), https://doi.org/10.1186/s40561-023-00244-y.