JOURNAL OF RESEARCH, INNOVATION, AND STRATEGIES FOR EDUCATION
VOL.3, NO.4, PP. 230-247, AUGUST 2026
https://doi.org/10.70148/rise.v3i4.14

PERTUBUHAN TEKNOLOGI DAN INOVASI PENDIDIKAN MALAYSIA

How Grade 11 Students Match Audio to Function Graphs: Distractor Patterns and Error Pathways

Samuel John Parreño [1,2,3*] **and Alleander Kate Gicale** [2,4]

[1] Department of Teacher Education, University of Mindanao, Digos City, Davao del Sur, Philippines
[2] Senior High School Department, Padada National High School, Padada, Davao del Sur, Philippines
[3] Faculty of Education, University of the Philippines Open University, Los Baños, Laguna, Philippines
[4] College of Graduate and Professional Education, Davao del Sur State College, Digos City, Philippines

ABSTRACT

Multiple-choice responses can reveal which graph features guide students' choices, not only whether an answer is correct. This study examined distractor patterns and pretest-to-posttest error pathways in audio-to-graph matching among 194 Grade 11 students from seven classes in a Philippine public secondary school. Parallel nine-item forms used level-coded audio, in which pitch tracked graph value, and rate-coded audio, in which pitch tracked local change. Responses were classified by correctness, graph family, mapping mismatch, and mathematical error category. Mean scores increased from 2.12 (SD = 1.32) to 3.64 (SD = 1.72) out of 9, $W = 1972$, $p < .001$, $r = .601$. Level-coded tasks showed marked movement from rate-coded lures to correct responses. Rate-coded tasks retained broader patterns of graph-level fallback, constant-versus-changing-rate confusion, concavity confusion, and nonlinear-rate confusion. Across 1,746 paired item attempts, 31.0% moved from incorrect to correct and 28.8% shifted between different incorrect categories. The study contributes a diagnostic pathway approach that identifies which distinctions in graph level, slope, concavity, and changing rate stabilized after instruction and which remained unsettled.

ARTICLE HISTORY

Received 24 April 2026
Revised 21 June 2026
Accepted 22 June 2026
Online First 22 June 2026
Published 06 August 2026

KEYWORDS

audio-to-graph matching; graph interpretation; distractor analysis; error pathways; mathematics assessment

1. Introduction

Interpreting function graphs is a central but persistently difficult part of secondary mathematics. Learners must coordinate graph level, direction of change, steepness, and changing rate, while also recognizing that the same visual feature may carry different meanings across contexts. Reviews and empirical studies of graph interpretation show that students often rely on surface features, isolated points, or perceptually salient shapes rather than coordinated reasoning about quantities and relationships (Friel et al., 2001; Glazer, 2011; Leinhardt et al., 1990; Shah & Hoeffner, 2002). In dynamic situations, graph reasoning is especially demanding because students must interpret a graph as a representation of covarying quantities rather than as a static picture (Carlson et al., 2002; Johnson et al., 2024).

CONTACT Samuel John Parreño samueljohn_parreno@umindanao.edu.ph Department of Teacher Education, University of Mindanao, Digos City, Davao del Sur, Philippines ISSN : 3030-6582 (Printed), eISSN : 3030-5330 (Online). DOI 10.70148/rise.v3i4.14.

A major source of difficulty is the distinction between graph height and rate of change. Students may treat a high point as a fast-changing point, a falling pitch or line as a decreasing quantity, or a flat-looking region as conceptually identical to a constant rate. Similar confusions have been documented in mathematics and physics settings, including difficulties with slope, kinematics graphs, and graph interpretation across domains (Beichner, 1994; Ivanjek et al., 2016; McDermott et al., 1987; Planinic et al., 2013). Research on slope and rate of change further shows that students' meanings for slope may remain fragmented across algebra, pre-calculus, and calculus contexts (Byerley & Thompson, 2017; Moore-Russo & Nagle, 2024; Nagle et al., 2013).

Multiple representations can support this kind of learning, but their usefulness depends on how clearly learners are helped to coordinate what each representation encodes. Theories of semiotic representation and multimedia learning emphasize that students must translate, connect, and regulate information across representational systems rather than simply receive more formats (Ainsworth, 2006; Duval, 2006; Mayer, 2020). Recent evidence on multiple external representations in STEM likewise suggests that the number of representations is less important than the interpretive support provided for using them (Rexigel et al., 2024). Universal Design for Learning also frames multiple means of representation as a pathway for access, provided that the representation clarifies rather than obscures disciplinary meaning (CAST, 2024).

Audio-based graph representations extend this problem into a temporal and auditory channel. Sonification translates data relations into non-speech sound, and prior work has shown that it can support scientific and mathematical interpretation when mapping rules are explicit and listeners know what sound features represent (Dubus & Bresin, 2013; Grond & Hermann, 2012; Hermann et al., 2011; Ohshiro et al., 2021; Tomlinson & Walker, 2020). In audio-to-graph matching tasks, however, correct answers alone do not fully reveal how students understand the sound-graph relation. A learner who selects a wrong graph may be following a systematic but inappropriate cue, such as choosing a graph with the highest visual point when pitch actually represents steepness, or choosing a graph with changing curvature when the sound represents constant slope.

For that reason, distractor analysis is an appropriate lens for studying pitch-coded graph interpretation. Multiple-choice distractors are not merely incorrect options. Well-designed distractors can serve as diagnostic indicators of competing interpretations, partial knowledge, and recurring misconceptions (Attali & Fraenkel, 2000; DiBattista & Kurzawa, 2011; Gierl et al., 2017; Haladyna et al., 2002; Rodriguez, 2005; Tarrant & Ware, 2010; Tarrant et al., 2009). In mathematics assessment, this diagnostic function is particularly important because a correct/incorrect score can conceal whether students are confusing graph level with rate, direction with steepness, or constant value with constant slope.

A descriptive error-pathway design is well suited to this problem because it preserves the identity of wrong responses across parallel tasks. A total-score comparison establishes whether performance changed, while the pathway analysis locates the mathematical feature attached to each response and follows that feature from pretest to posttest. A shift between two incorrect categories can show that students redirected attention even when the item remained incorrect. Binary scores collapse that movement into a single zero. Recent mathematics diagnostic research likewise shows that item and option behavior can improve the information produced by an assessment (Akveld & Kinnear, 2024).

The study examined distractor patterns and response pathways in audio-to-graph tasks for secondary function interpretation. Four questions guided the analysis: (1) Which distractors and error categories were most frequently selected before and after instruction? (2) How did error profiles differ between level-coded and rate-coded tasks? (3) How did responses move among correct choices, persistent wrong categories, and shifted wrong categories? (4) Which dominant pathways reveal continuing difficulty with graph level, direction, constant rate, concavity, diminishing change, and nonlinear rate?

2. Conceptual Framework

The framework for the study joins four lines of work: graph interpretation, multiple representation, sonification, and distractor analysis. First, graph interpretation is treated as covariational reasoning. Students need to coordinate the values of quantities and how those values change across an interval (Carlson et al., 2002; Johnson et al., 2024). This view is compatible with classic work showing that graphs are difficult because they compress multiple meanings, including point values, global shape, local slope, and contextual change (Friel et al., 2001; Leinhardt et al., 1990; Shah & Hoeffner, 2002).

Second, the study treats slope and rate of change as conceptually fragile meanings. Kinematics and cross-context graph studies show that learners frequently confuse slope, height, direction, and area or may transfer an interpretation that is valid in one graphing context to another where it is not valid (Beichner, 1994; Ivanjek et al., 2016; McDermott et al., 1987; Planinic et al., 2013). Studies of students' and teachers' meanings for slope also show that rate of change can be understood as a formula, visual steepness, ratio, or measure, and these meanings do not always cohere during problem solving (Byerley & Thompson, 2017; Moore-Russo & Nagle, 2024; Nagle et al., 2013). In pitch-coded tasks, these difficulties may surface as graph-level fallback when students hear pitch as height even when it encodes local change.

Third, the study assumes that audio representations must be interpreted as part of a larger representational system. Multiple representations support learning when students understand what each representation contributes and how it links with others (Ainsworth, 2006; Duval, 2006; Mayer, 2020; Rexigel et al., 2024). In sonification, this principle appears as the need for clear mapping between data features and sound features (Dubus & Bresin, 2013; Grond & Hermann, 2012; Hermann et al., 2011). Audio-to-graph matching therefore requires more than hearing a sound. It requires identifying the active mapping, aligning the sound with the left-to-right domain scan, and selecting a graph whose visual structure corresponds to the encoded meaning.

Fourth, distractors are treated as evidence about competing interpretations. Multiple-choice assessment research emphasizes that distractors should be analyzed because they may identify non-functioning options, common misconceptions, or partial knowledge states (Attali & Fraenkel, 2000; DiBattista & Kurzawa, 2011; Gierl et al., 2017; Haladyna et al., 2002; Rodriguez, 2005; Tarrant & Ware, 2010; Tarrant et al., 2009). A diagnostic assessment perspective also values evidence that helps teachers decide what feedback or instructional adjustment is needed (Black & Wiliam, 1998; Frederiksen & Collins, 1989; Sadler, 1989). In this study, distractor selections are coded not only by letter but also by their mathematical interpretation, allowing wrong responses to be treated as error pathways rather than as undifferentiated failure.

Figure 1 connects these perspectives to the study's analytical sequence. The audio mapping establishes the mathematical feature to track. Students may then attend to graph level, direction, steepness, constancy of rate, concavity, or nonlinear change. The selected option provides evidence about the interpretation that competed with the target, and the paired responses distinguish stable correctness, resolved errors, persistent errors, shifted errors, and movement from correct to incorrect.

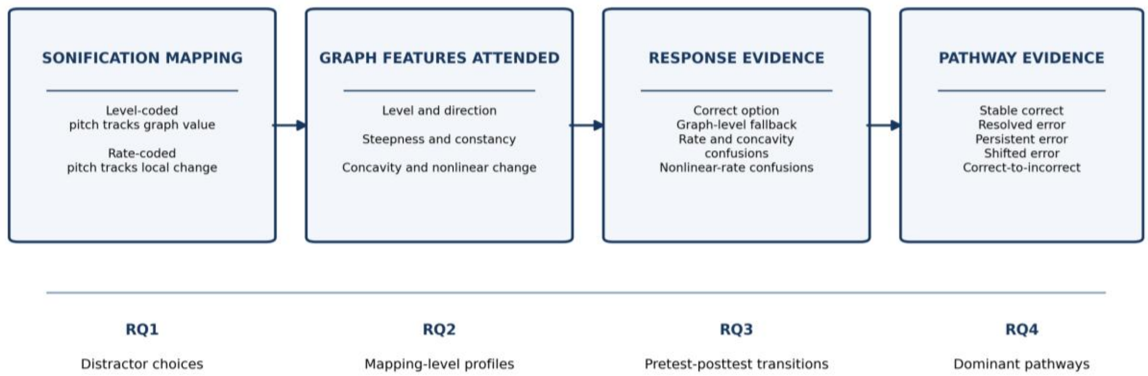

Figure 1. Operational framework linking sonification mapping, attended graph features, response evidence, and pretest-to-posttest pathways.

3. Method

3.1. Research design and setting

The study used a descriptive pretest-posttest error-pathway design in Grade 11 General Mathematics at a public secondary school in Padada, Davao del Sur, Philippines. Seven intact classes completed the same classroom sequence. The design supported two complementary analyses: a paired total-score analysis at the student level and a diagnostic analysis of option-level movements across construct-paired items. The first summarized short-term performance change; the second identified the graph interpretation attached to each selected option.

3.2. Participants and contextual profile

The analytic sample comprised 194 students with complete paired records on both nine-item forms. Study codes linked pretest and posttest responses, and student names were not included in the analytic files. Table 1 reports the contextual and demographic information available for the article. Sex categories were recorded in classroom documentation, but aggregate counts were not retained in the response-transition file. Age, academic strand or track, and prior mathematics achievement were not collected, so no retrospective distributions were constructed.

Table 1: Participant and study context.

Characteristic	Available information
Educational level	Grade 11
Course	General Mathematics
Setting	Public secondary school
Location	Padada, Davao del Sur, Philippines
Participating classes	Seven intact classes
Complete paired records	194 students
Sex	Categories recorded
Age, strand/track, prior achievement	Not collected

3.3. Instructional sequence

Each class followed a common pretest-instruction-posttest sequence. Students first completed a nine-item audio-to-graph pretest. A brief concept launch introduced two mappings: value-to-pitch, where pitch followed graph height, and slope-to-pitch, where pitch followed signed local change. The teacher then led two predict-listen-discuss cycles. During the value-to-pitch cycle, students predicted whether pitch would rise, fall, or remain constant, listened to the clip, and explained how the sound followed graph height. During the slope-to-pitch cycle, students predicted where change would be fastest, listened for high and low pitch regions, and connected those regions to steepness and the sign of slope. The teacher explicitly contrasted a graph's highest point with its steepest point. Students completed the parallel posttest immediately after the guided activities.

The materials used a fixed A-I graph panel, a projector, and classroom speakers, with locally stored audio available when internet access was unreliable. Clips used a sine-wave timbre, a 220-880 Hz pitch range, 40 left-to-right scan steps, and an approximate duration of nine seconds. Each clip was played once, with one replay permitted when most students requested it. These controls kept the auditory presentation consistent across classes.

3.4. Assessment forms and item blueprint

The pretest and posttest contained nine construct-paired items with the same response format and parallel mathematical targets. Three pairs were level-coded and six were rate-coded. Level-coded items used pitch to represent graph value, while rate-coded items used pitch to represent local change or steepness. The forms sampled constant, linear, exponential, quadratic, logarithmic, square-root, cubic, and logistic patterns. Table 2 summarizes the item pairing and answer keys.

Table 2: Parallel item blueprint and correct options.

Pair	Mapping	Pretest target (key)	Posttest target (key)
1	Level	positive linear level (B)	negative linear level (B)
2	Rate	constant slope (I)	constant slope (I)
3	Rate	concavity up / increasing rate (D)	concavity up / increasing rate (D)
4	Rate	diminishing negative square-root pattern (A)	diminishing logarithmic change (A)
5	Level	exponential decay level (H)	exponential growth level (H)
6	Rate	diminishing logarithmic change (F)	cubic changing-rate segment (F)
7	Level	exponential growth level (E)	constant level (E)
8	Rate	peak-rate logistic pattern (G)	diminishing negative square-root pattern (G)
9	Rate	cubic increasing-rate segment (C)	peak-rate logistic pattern (C)

Note. Level-coded means pitch tracked graph value; rate-coded means pitch tracked local change or steepness. Item targets and correct options were linked from the assessment specifications before analysis.

3.5. Instrument evidence and reliability

Content evidence came from the shared nine-target blueprint, curriculum alignment, and item review by the two mathematics educators who developed the lesson. The review checked the intended audio mapping, function family, graph feature, answer key, and plausibility of the A-I options. The forms used parallel constructs and standardized audio parameters. They were not formally equated, and no independent pilot panel was conducted; item-level findings are therefore reported with the specific pretest and posttest targets.

Internal consistency was KR-20 = .22 for the pretest and .47 for the posttest. These estimates are low for broad achievement inference and reflect a short form spanning several graph features and function families. The analysis therefore combines total-score change, item-level tests, distractor profiles, and response pathways instead of relying on a single coefficient.

3.6. Distractor coding and quality control

Each response was first classified as correct or incorrect. Incorrect options were then assigned to a mathematical category according to the relation between the selected graph and the target. The categories were graph-level fallback in a rate-coded item; constant-rate versus changing-rate confusion; concavity or diminishing-rate confusion; peak-rate or nonlinear-rate confusion; rate-coded lure in a level-coded item; level-direction confusion; and within-level shape confusion.

The item-option crosswalk was checked against the graph panel, answer key, audio-mapping rule, and item specification before frequencies were computed. A fixed lookup table then assigned every A-I response to a category. Individual student responses did not require repeated subjective coding, so response-level inter-rater agreement was not calculated. The quality-control step focused on the crosswalk that generated all 1,746 item-level labels.

3.7. Data handling and analysis

The raw file contained 200 records. Six records lacked at least one pretest response and could not support a complete paired pathway, leaving 194 complete pairs. The excluded records represented 3.0% of the raw file. Imputation was not used because an inferred item response would create a transition that was not observed.

Pretest, posttest, and gain scores were summarized with means, standard deviations, medians, and interquartile ranges. Gain scores departed from normality in the Shapiro-Wilk test ($p < .001$), so the paired total-score comparison used the Wilcoxon signed-rank test. The Hodges-Lehmann estimate, 95% confidence interval, and matched-pairs effect size r described the size and precision of change. Item-level changes in correct-response proportions were examined with paired McNemar tests and Holm adjustment across the nine item pairs.

The diagnostic analysis computed option frequencies, dominant distractors, and error-category profiles by phase and audio-mapping type. Each paired item response was classified as correct to correct, correct to incorrect, incorrect to correct, incorrect to incorrect within the same category, or incorrect to incorrect with a shifted category. Pathway percentages remained descriptive because the 1,746 item attempts were nested within 194 students and were used to map response structure, not as independent observations for a treatment-effect test.

3.8. Ethical considerations

The study was conducted under school and Schools Division classroom research procedures for minimal-risk educational studies. Written parent or guardian consent and learner assent were obtained before records were included. Participation was voluntary and carried no grade penalty. Students could decline or withdraw without penalty. Study codes replaced names in the analytic files, and only de-identified summaries are reported. The approval process was documented through the applicable school and division records; no numerical IRB identifier was assigned.

4. Results

Results are organized around overall paired performance, distractor selections, mapping-level error profiles, response transitions, dominant pathways, and item-family patterns. Percentages are based on 194 complete paired records unless otherwise stated.

4.1. Overall paired performance

Overall scores increased from M = 2.12 (SD = 1.32) on the pretest to M = 3.64 (SD = 1.72) on the posttest. The mean gain was 1.53 points out of nine (SD = 2.05), and the median rose from 2 (IQR = 1-3) to 4 (IQR = 2-5). The Wilcoxon signed-rank test showed a significant paired difference, W = 1972, $p < .001$, with r = .601. The Hodges-Lehmann estimate was a one-point increase, 95% CI [1, 2]. Holm-adjusted McNemar tests identified significant increases on five item pairs and significant declines on two; two pairs showed no significant change.

4.2. Distractor selections by item and phase

Table 3 shows that many item-phase pairs were dominated by one mathematically plausible wrong option. Before instruction, the level-coded Item 1 was led by a cubic increasing-rate lure (45.4%), and Item 7 was led by a positive linear-level distractor (38.1%). After instruction, dominant distractors declined sharply on the level-coded items. Several rate-coded items remained dispersed across constant-rate, concavity, diminishing-rate, and nonlinear-rate alternatives; Posttest Item 4 had only 5.2% correct responses.

Table 3: Dominant pretest and posttest distractor frequencies by item.

Phase	Item	Correct option	Correct %	Dominant distractor	n	%	Error category
Pre	1	B	17.5	C (cubic increasing rate)	88	45.4	Rate-coded lure in level-coded item
Pre	2	I	44.3	B (linear positive)	39	20.1	Graph-level fallback in rate-coded item
Pre	3	D	14.9	E (exponential growth)	56	28.9	Graph-level fallback in rate-coded item
Pre	4	A	10.3	G (logistic peak rate)	77	39.7	Peak-rate or nonlinear-rate confusion
Pre	5	H	30.4	A (square-root concavity down)	40	20.6	Rate-coded lure in level-coded item
Pre	6	F	16.5	I (constant slope linear)	31	16.0	Constant-rate versus changing-rate confusion
Pre	7	E	16.5	B (linear positive)	74	38.1	Level-direction confusion
Pre	8	G	55.7	F (logarithmic	21	10.8	Concavity or diminishing-

				diminishing change)			rate confusion
Pre	9	C	5.7	H (exponential decay)	47	24.2	Graph-level fallback in rate-coded item
Post	1	B	71.1	A (logarithmic diminishing change)	15	7.7	Rate-coded lure in level-coded item
Post	2	I	36.6	E (constant level)	32	16.5	Graph-level fallback in rate-coded item
Post	3	D	41.8	I (constant slope linear)	46	23.7	Constant-rate versus changing-rate confusion
Post	4	A	5.2	C (logistic peak rate)	46	23.7	Peak-rate or nonlinear-rate confusion
Post	5	H	59.3	B (linear negative)	25	12.9	Level-direction confusion
Post	6	F	6.7	A (logarithmic diminishing change)	77	39.7	Concavity or diminishing-rate confusion
Post	7	E	65.5	H (exponential growth)	30	15.5	Within-level shape confusion
Post	8	G	29.4	F (cubic changing rate)	52	26.8	Peak-rate or nonlinear-rate confusion
Post	9	C	49.0	F (cubic changing rate)	37	19.1	Peak-rate or nonlinear-rate confusion

Note. The table reports the most frequently selected incorrect option for each item-phase pair. Correct % is the percentage selecting the correct option.

Figure 2 integrates correct responses and error categories across all item-phase pairs. The level-coded rows shift toward the correct-response column, while the rate-coded rows remain distributed across several interpretations of local change. The figure therefore shows both error reduction and error redistribution.

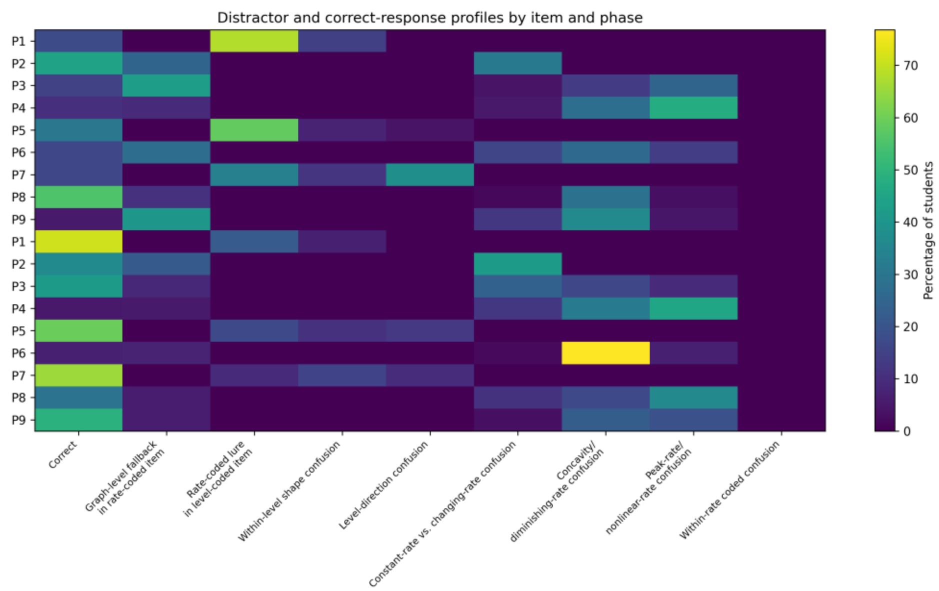

Figure 2. Heatmap of distractor selections and correct-response profiles by item and test phase.

4.3. Error-category profiles by audio-mapping type

Table 4 summarizes wrong responses by phase and audio-mapping type. Rate-coded lures accounted for 67.8% of wrong level-coded responses at pretest and 46.5% at posttest. In rate-coded tasks, graph-level fallback decreased from 34.4% to 13.0% of wrong responses, while concavity or diminishing-rate confusion reached 38.1% and peak-rate or nonlinear-rate confusion reached 27.0% at posttest. The profile moved away from broad graph-level fallback but remained divided among finer distinctions in local change.

Table 4: Error-category frequencies by mapping type.

Phase	Task Mapping	Error Category	n	% within wrong responses
Post	Level-coded	Rate-coded lure in level-coded item	94	46.5
Post	Level-coded	Within-level shape confusion	64	31.7
Post	Level-coded	Level-direction confusion	44	21.8
Post	Rate-coded	Concavity or diminishing-rate confusion	319	38.1
Post	Rate-coded	Peak-rate or nonlinear-rate confusion	226	27.0
Post	Rate-coded	Constant-rate versus changing-rate confusion	183	21.9
Post	Rate-coded	Graph-level fallback in rate-coded item	109	13.0
Pre	Level-coded	Rate-coded lure in level-coded item	310	67.8
Pre	Level-coded	Level-direction confusion	82	17.9
Pre	Level-coded	Within-level shape confusion	65	14.2

Pre	Rate-coded	Graph-level fallback in rate-coded item	302	34.4
Pre	Rate-coded	Concavity or diminishing-rate confusion	258	29.4
Pre	Rate-coded	Peak-rate or nonlinear-rate confusion	182	20.7
Pre	Rate-coded	Constant-rate versus changing-rate confusion	136	15.5

Note. Percentages are computed within wrong responses for each phase and mapping type, not within all item attempts.

Figure 3 displays the mapping-level contrast. Level-coded errors are concentrated in a smaller set of alternatives, whereas rate-coded errors form a broader cluster involving graph level, constancy of rate, concavity, and nonlinear change.

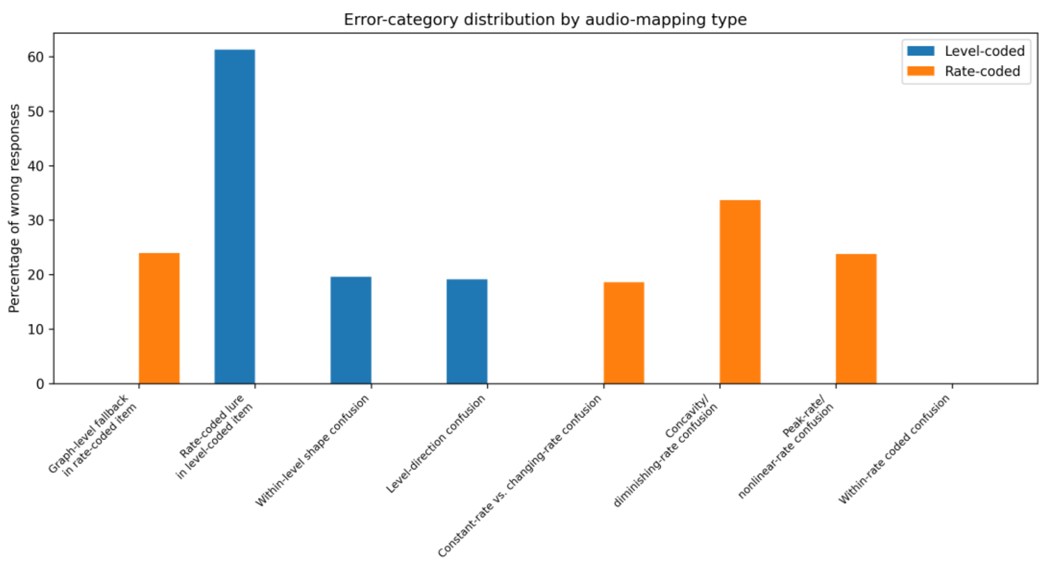

Figure 3. Error-category distribution by audio-mapping type.

4.4. Pretest-to-posttest response transitions

Table 5 reports 1,746 paired item transitions. Incorrect-to-correct movement accounted for 31.0%, while 28.8% shifted from one incorrect category to another and 16.7% remained within the same incorrect category. The mapping contrast was marked: 51.9% of level-coded transitions moved from incorrect to correct, whereas 36.1% of rate-coded transitions shifted between different incorrect categories.

Table 5: Pretest-to-posttest transition matrix

Transition status	Level-coded	Rate-coded	Total
Correct → Correct	78 (13.4%)	88 (7.6%)	166 (9.5%)
Correct → Incorrect	47 (8.1%)	198 (17.0%)	245 (14.0%)
Incorrect → Correct	302 (51.9%)	239 (20.5%)	541 (31.0%)

Incorrect → Incorrect, same error category	73 (12.5%)	219 (18.8%)	292 (16.7%)
Incorrect → Incorrect, shifted category	82 (14.1%)	420 (36.1%)	502 (28.8%)

Note. Percentages within the Level-coded and Rate-coded columns use the number of paired attempts for that mapping type as denominator. Total percentages use all paired item attempts as denominator.

Figure 4 shows that incorrect-to-correct movement was the largest single pathway, but shifted-category incorrect responses were nearly as frequent. This closeness establishes why pathway evidence adds information that a binary pretest-posttest score cannot retain.

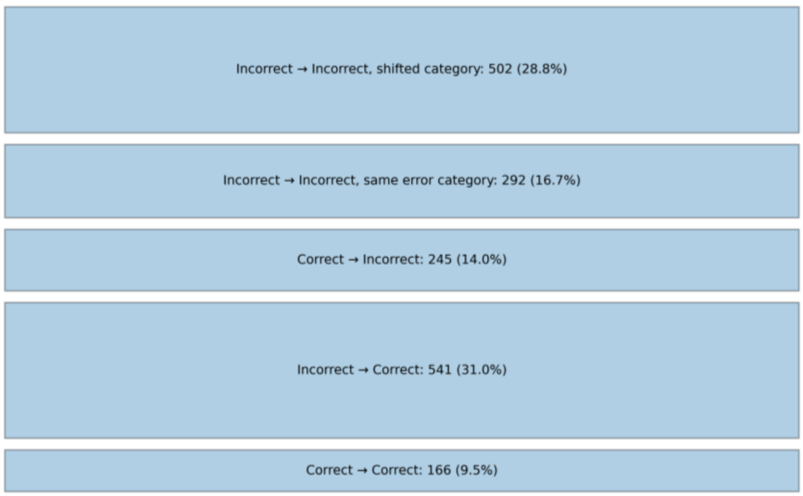

Figure 4. Network of incorrect-to-incorrect error-category movements from pretest to posttest.

4.5. Dominant error pathways

Table 6 identifies the largest item-level pathways. In Level-coded Item 1, 48.5% of students moved from a rate-coded lure to the correct response. Level-coded Items 5 and 7 also showed prominent movements to correctness. Rate-coded Items 4 and 6 retained or exchanged nonlinear-rate and concavity-related errors, indicating that several students changed the feature they attended to without stabilizing the intended relation.

Table 6: Dominant error pathways and mathematical interpretation

mapping	item	pre_error_category	post_error_category	n	% of item N
Level-coded	1	Rate-coded lure in level-coded item	Correct	94	48.5
Level-coded	5	Rate-coded lure in level-coded item	Correct	62	32.0
Level-coded	7	Level-direction confusion	Correct	57	29.4
Rate-coded	4	Peak-rate or nonlinear-rate confusion	Peak-rate or nonlinear-rate confusion	56	28.9

Level-coded	7	Rate-coded lure in level-coded item	Correct	42	21.6
Rate-coded	6	Concavity or diminishing-rate confusion	Concavity or diminishing-rate confusion	41	21.1
Rate-coded	9	Graph-level fallback in rate-coded item	Correct	40	20.6
Rate-coded	6	Graph-level fallback in rate-coded item	Concavity or diminishing-rate confusion	35	18.0
Rate-coded	9	Concavity or diminishing-rate confusion	Correct	35	18.0
Rate-coded	3	Graph-level fallback in rate-coded item	Correct	34	17.5
Rate-coded	2	Constant-rate versus changing-rate confusion	Constant-rate versus changing-rate confusion	27	13.9
Level-coded	1	Rate-coded lure in level-coded item	Rate-coded lure in level-coded item	27	13.9

Note. The table shows the largest pretest-to-posttest movements involving at least one incorrect response category. Item-level percentages use 194 students as the denominator.

Figure 5 maps the main incorrect-to-incorrect category movements. The strongest persistence appears within concavity or diminishing-rate confusion. Connections from graph-level fallback to concavity-related errors and between nonlinear-rate and concavity-related errors show a connected field of rate interpretations, not isolated distractors.

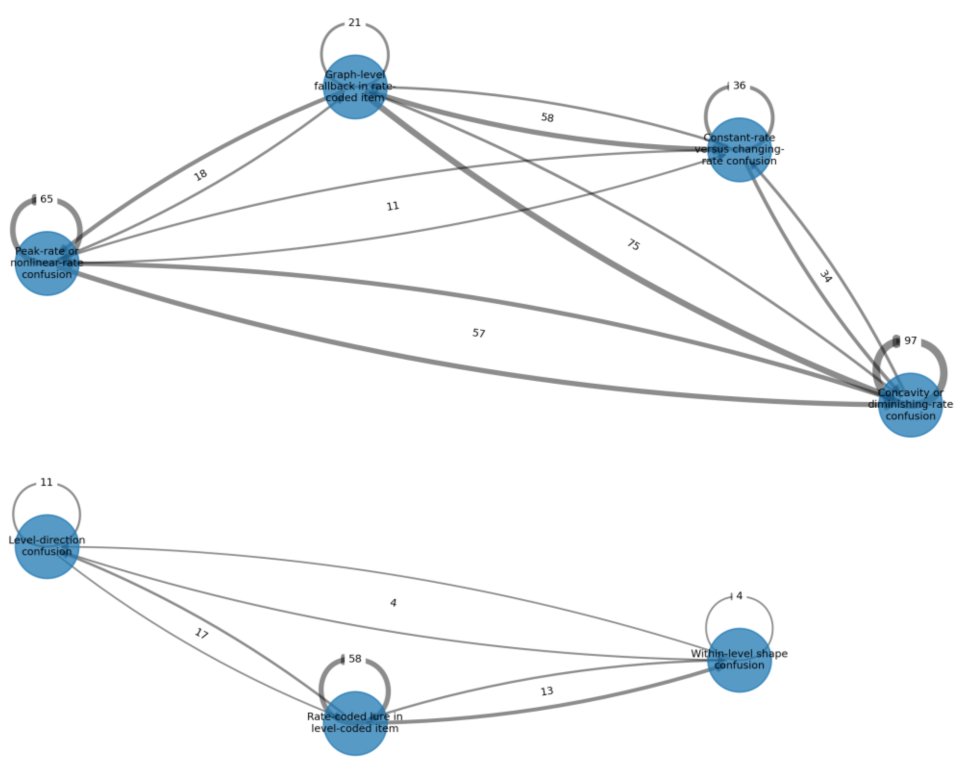

Figure 5. Response-transition pathway distribution from pretest to posttest.

4.6. Item-family summary of persistent and reduced errors

Table 7 summarizes item-family patterns. All three level-coded pairs showed reduced dominant errors. Four of the six rate-coded pairs remained persistent or increased, especially the pairs involving diminishing change, cubic change, and square-root-type change. Rate-coded Item 9 showed the clearest reduction, although its remaining errors still centered on concavity or diminishing-rate distinctions.

Table 7: Summary of persistent and reduced error patterns by item family

Item	Mapping	Pretest target	Posttest target	Pretest accuracy %	Posttest accuracy %	Dominant pretest error	Dominant posttest error	Pattern
1	Level-coded	positive linear level	negative linear level	17.5	71.1	Rate-coded lure in level-coded item	Rate-coded lure in level-coded item	Reduced
2	Rate-coded	constant slope	constant slope	44.3	36.6	Constant-rate versus changing-rate confusion	Constant-rate versus changing-rate confusion	Persistent/increased
3	Rate-coded	concavity up / increasing rate	concavity up / increasing rate	14.9	41.8	Graph-level fallback in rate-coded item	Constant-rate versus changing-rate confusion	Reduced
4	Rate-coded	diminishing negative square-root pattern	diminishing logarithmic change	10.3	5.2	Peak-rate or nonlinear-rate confusion	Peak-rate or nonlinear-rate confusion	Persistent/increased
5	Level-coded	exponential decay level	exponential growth level	30.4	59.3	Rate-coded lure in level-coded item	Rate-coded lure in level-coded item	Reduced
6	Rate-coded	diminishing logarithmic change	cubic changing-rate segment	16.5	6.7	Graph-level fallback in rate-coded item	Concavity or diminishing-rate confusion	Persistent/increased
7	Level-coded	exponential growth level	constant level	16.5	65.5	Level-direction confusion	Within-level shape confusion	Reduced
8	Rate-coded	peak-rate logistic pattern	diminishing negative square-root pattern	55.7	29.4	Concavity or diminishing-rate confusion	Peak-rate or nonlinear-rate confusion	Persistent/increasing

9	Rate-coded	cubic increasing-rate segment	peak-rate logistic pattern	5.7	49.0	Graph-level fallback in rate-coded item	Concavity or diminishing-rate confusion	Reduced

Note. Pattern is based on change in accuracy and dominant wrong-response category. It is descriptive and intended for instructional diagnosis.

Figure 6 highlights where a single distractor captured a large share of responses. The dominant bars contracted across the level-coded items after instruction but remained prominent for difficult rate-coded targets. These concentrations identify specific options for item revision and focused reteaching.

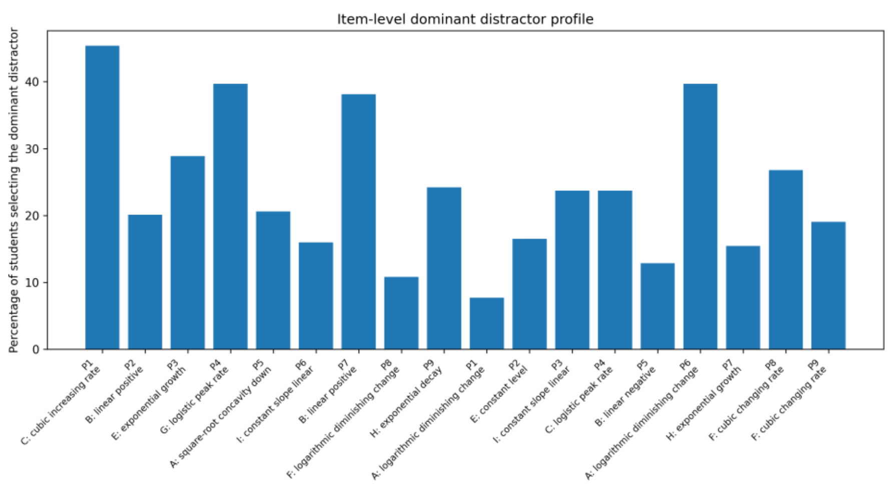

Figure 6. Dominant distractor profile by item and test phase.

5. Discussion

The findings show that wrong responses carried structured diagnostic information. Level-coded errors frequently reflected attention to rate-like shapes even though pitch followed graph value, and those lures often resolved after the explicit value-to-pitch cycle. Rate-coded tasks produced a wider and more persistent field of alternatives involving graph level, constant rate, concavity, diminishing change, and nonlinear rate. This contrast is consistent with research showing that slope and rate meanings remain fragmented across representations and contexts (Byerley & Thompson, 2017; Ivanjek et al., 2016; Moore-Russo & Nagle, 2024).

The main theoretical contribution is a response-state account of representational learning. Binary scoring records correctness at each time point. Error pathways preserve the mathematical category attached to each wrong choice and show whether an interpretation persisted, resolved, or changed form. The 28.8% shifted-category pathway indicates that a substantial share of post-instruction movement occurred within the incorrect-response space. Movement from graph-level fallback to concavity confusion, for example, places attention closer to local change while leaving the finer distinction unresolved.

This response-state account extends diagnostic assessment and representational coordination theory. Students did not simply possess or lack the correct audio rule. Their choices show attention moving among level, direction, steepness, constancy, and curvature. Rate-coded audio placed several related distinctions into competition, which helps explain the mobility among error categories. The pattern accords with evidence linking graph selections to the quality of graph reasoning (Johnson et al., 2024) and with the view that multiple representations require active coordination (Ainsworth, 2006; Rexigel et al., 2024).

Several mechanisms may have contributed to the pathways. The forms were construct-paired but not identical, so item difficulty and the visual attractiveness of particular graphs may have influenced category movement. The posttest followed the lesson immediately, leaving short-term cue retention and practice effects possible. Classroom acoustics, replay requests, and attention during whole-class listening may also have varied across the seven classes. The low KR-20 values indicate limited internal consistency for the brief heterogeneous forms. Selected distractors provide indirect evidence about reasoning; learner explanations or think-aloud data are needed to confirm the meaning attached to each choice.

Shifted-category pathways can therefore mark partial conceptual reorganization. Movement away from graph-level fallback toward a rate-based category may indicate attention to the relevant quantity, even when the distinction among constant, diminishing, and nonlinear rate remains unstable. Such movement should be checked through explanations, delayed assessment, or transfer tasks before it is treated as durable understanding.

The pathway framework also strengthens formative use of multiple-choice evidence. Option-level patterns can identify the next contrast to teach, the distractor that needs redesign, and the distinction that remains fragile. This use aligns with formative-assessment principles that connect evidence to a specific instructional response (Black & Wiliam, 1998; Frederiksen & Collins, 1989; Sadler, 1989).

5.1. Limitations

Several limitations define the reach of the findings. The study involved one public secondary school, seven intact classes, and an immediate posttest, so the results do not establish delayed retention or transfer. The nine-item forms sampled several function families, were not formally equated, and produced low KR-20 estimates. Parallel-form difficulty and distractor attractiveness may therefore account for part of the item-level movement. The category scheme was grounded in graph-option meanings and fixed before frequency analysis, but it has not yet been independently validated through a larger expert panel or learner think-aloud study.

Six of 200 records were excluded because a complete paired pathway could not be formed. Complete-case analysis was appropriate for observed transition sequences, but it may underrepresent students with incomplete responses. Future studies should reduce missingness prospectively and may use sensitivity analyses for score-based outcomes; imputation is less suitable for pathway analysis because it would introduce unobserved option transitions. Aggregate sex counts were not retained in the analytic file, and age, academic strand or track, and prior mathematics achievement were not collected. Replication across schools and learner groups should include these contextual variables.

5.2. Implications for Mathematics Teaching and Assessment

The pathway profiles support contrastive classroom activities. A teacher can place a graph’s highest point beside its steepest point and ask students to predict how each location would sound under value-to-pitch and slope-to-pitch mappings. A constant function can then be paired with a nonhorizontal line to contrast constant value with constant slope. For increasing curves that flatten, students can mark intervals where the function rises while the rate decreases, listen to the clip, and explain why falling pitch does not always mean falling graph value.

For assessment, distractors should be designed around specific mathematical contrasts and reviewed after use. Option frequencies can show whether students are relying on graph level, constant rate, concavity, or a salient nonlinear shape. A teacher can use the pretest-posttest pathway table as a compact formative record: resolved errors suggest that the contrast has become clearer; persistent errors call for reteaching; shifted errors identify the next distinction that needs attention.

Audio should remain synchronized with visible graphs, explicit mapping labels, and opportunities for students to verbalize or annotate their reasoning. It can widen the representational field, but it does not replace visual, symbolic, or verbal access. Consistent playback, clear left-to-right scanning, and low-noise alternatives are essential when whole-class sound is used.

6. Conclusion

Distractor and error-pathway analysis revealed which mathematical features guided students' audio-to-graph choices and how those choices changed after instruction. Level-coded tasks showed broad movement toward correctness, while rate-coded tasks retained a connected set of confusions involving graph level, constancy, concavity, diminishing change, and nonlinear rate. The near balance between incorrect-to-correct movement (31.0%) and shifted-category incorrect movement (28.8%) shows why total scores alone provide an incomplete account of short-term learning.

The study contributes a practical diagnostic framework for converting selected options into evidence about response states. In formative assessment, the framework can connect a dominant distractor or pathway to a specific contrastive activity, feedback prompt, or item revision. Future research should independently validate the category scheme, use delayed and transfer assessments, collect concurrent learner explanations, examine subgroup patterns, and test the framework across schools, graph types, and audio mappings.

Declaration

Competing Interests
None.

Ethical Approval
The study was conducted under school and Schools Division classroom research procedures for minimal-risk educational studies. Written parent or guardian consent and learner assent were obtained before records were included. Participation was voluntary and carried no grade penalty. Students could decline or withdraw without penalty. Study codes replaced names in the analytic files, and results are reported only in de-identified form. The approval process was documented through the school and division records; no numerical IRB identifier was assigned.

Author's Contribution
Author1: Conceptualization, Formal analysis, Validation, Visualization, Writing – original draft, Supervision, Investigation
Author2: Methodology, Resources, Writing – review and editing

Data availability
Not applicable

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Akveld, M., & Kinnear, G. (2024). Improving mathematics diagnostic tests using item analysis. International Journal of Mathematical Education in Science and Technology, 55(10), 2478-2505. https://doi.org/10.1080/0020739X.2023.2167132

Attali, Y., & Fraenkel, T. (2000). The point-biserial as a discrimination index for distractors in multiple-choice items: Deficiencies in usage and an alternative. Journal of Educational Measurement, 37(1), 77–86. https://doi.org/10.1111/j.1745-3984.2000.tb01077.x

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. American Journal of Physics, 62(8), 750–762. https://doi.org/10.1119/1.17449

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

Byerley, C., & Thompson, P. W. (2017). Secondary mathematics teachers' meanings for measure, slope, and rate of change. The Journal of Mathematical Behavior, 48, 168–193. https://doi.org/10.1016/j.jmathb.2017.09.003

Carlson, M. P., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. Journal for Research in Mathematics Education, 33(5), 352–378. https://doi.org/10.2307/4149958

CAST. (2024). Universal Design for Learning guidelines version 3.0. https://udlguidelines.cast.org/

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. The Canadian Journal for the Scholarship of Teaching and Learning, 2(2), Article 4. https://doi.org/10.5206/cjsotl-rcacea.2011.2.4

Downing, S. M., & Haladyna, T. M. (Eds.). (2006). Handbook of test development. Lawrence Erlbaum Associates.

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. PLOS ONE, 8(12), e82491. https://doi.org/10.1371/journal.pone.0082491

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. Educational Studies in Mathematics, 61, 103–131. https://doi.org/10.1007/s10649-006-0400-z

Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Prentice Hall.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18(9), 27–32. https://doi.org/10.3102/0013189X018009027

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. Journal for Research in Mathematics Education, 32(2), 124–158. https://doi.org/10.2307/749671

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. Review of Educational Research, 87(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. Studies in Science Education, 47(2), 183–210. https://doi.org/10.1080/03057267.2011.605307

Grond, F., & Hermann, T. (2012). Aesthetic strategies in sonification. AI & Society, 27(2), 213–222. https://doi.org/10.1007/s00146-011-0341-7

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5

Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds.). (2011). The Sonification Handbook. Logos Verlag.

Ivanjek, L., Susac, A., Planinic, M., Andrasevic, A., & Milin-Sipus, Z. (2016). Student reasoning about graphs in different contexts. Physical Review Physics Education Research, 12(1), 010106. https://doi.org/10.1103/PhysRevPhysEducRes.12.010106

Johnson, H. L., Donovan, C., Knurek, R., Whitmore, K. A., & Bechtold, L. (2024). Proposing and testing a model relating students' graph selection and graph reasoning for dynamic situations. Educational Studies in Mathematics. https://doi.org/10.1007/s10649-024-10299-4

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. Review of Educational Research, 60(1), 1–64. https://doi.org/10.3102/00346543060001001

Mayer, R. E. (2020). Multimedia learning (3rd ed.). Cambridge University Press.
McDermott, L. C., Rosenquist, M. L., & van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. American Journal of Physics, 55(6), 503–513. https://doi.org/10.1119/1.15104
Moore-Russo, D., & Nagle, C. (2024). Connecting representations and ways of thinking about slope from algebra to calculus. The Mathematics Enthusiast, 21(3), 569–602. https://doi.org/10.54870/1551-3440.1643
Nagle, C., Moore-Russo, D., Viglietti, J., & Martin, K. (2013). Calculus students' and instructors' conceptualizations of slope: A comparison across academic levels. International Journal of Science and Mathematics Education, 11(6), 1491–1515. https://doi.org/10.1007/s10763-013-9411-2
Ohshiro, K., Hurst, A., & DuBois, L. (2021). Making math graphs more accessible in remote learning: Using sonification to introduce discontinuity in calculus. In Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Article 77, pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3441852.3476533
Planinic, M., Ivanjek, L., Susac, A., & Milin-Sipus, Z. (2013). Comparison of university students' understanding of graphs in different contexts. Physical Review Special Topics - Physics Education Research, 9(2), 020103. https://doi.org/10.1103/PhysRevSTPER.9.020103
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, Article 4. https://doi.org/10.1007/s10648-024-09958-y
Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educational Measurement: Issues and Practice, 24(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18, 119–144. https://doi.org/10.1007/BF00117714
Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. Educational Psychology Review, 14(1), 47–69. https://doi.org/10.1023/A:1013180410169
Tarrant, M., & Ware, J. (2010). A comparison of the psychometric properties of three- and four-option multiple-choice questions in nursing assessments. Nurse Education Today, 30(6), 539–543. https://doi.org/10.1016/j.nedt.2009.11.002
Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Medical Education, 9, Article 40. https://doi.org/10.1186/1472-6920-9-40
Tomlinson, B. J., & Walker, B. N. (2020). Auditory display in interactive science simulations: Description and sonification support interaction and enhance opportunities for learning. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376886