Wiley
Human Behavior and Emerging Technologies
Volume 2025, Article ID 8813532, 14 pages
https://doi.org/10.1155/hbe2/8813532

WILEY

Research Article

Cognitive Benefits of Employing Multiple AI Voices as Specialist Virtual Tutors in a Multimedia Learning Environment

Tze Wei Liew,[1] **Su-Mae Tan**,[2] **Tak Jie Chan**,[3] **Yang Tian**,[4] **and Faizan Ahmad**[5]

[1]*Centre for Interaction and Experience Design, Multimedia University, Cyberjaya, Selangor, Malaysia*
[2]*Faculty of Information Science & Technology, Multimedia University, Melaka, Malaysia*
[3]*Faculty of Business and Communications, INTI International University, Nilai, Negeri Sembilan, Malaysia*
[4]*Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang, Selangor, Malaysia*
[5]*Department of Computer Science, Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, UK*

Correspondence should be addressed to Tze Wei Liew; twliew@mmu.edu.my

Received 2 October 2024; Revised 15 February 2025; Accepted 7 August 2025

Academic Editor: Rosemary Fisher

Limited prior research provides some evidence of the cognitive and learning benefits of employing multiple pedagogical agents, each assigned to distinct knowledge bases, in a multimedia learning environment. However, follow-up studies and extensions of these findings remain scarce. To address this gap, we draw on multimedia learning and cognitive models to investigate the effects of using multiple AI voices as specialist virtual tutors for distinct programming algorithm subtopics on cognitive load and learning outcomes. A between-subjects experimental design was employed with first-year business undergraduates who had minimal programming knowledge. Participants engaged with a multimedia learning video, narrated either by a single AI voice or by three distinct AI voices, each assigned to a different subtopic. Cognitive load was measured via a survey, while learning outcomes were assessed using immediate and 2-week delayed posttests covering retention, near-transfer, and far-transfer tasks. Results indicated that participants in the multiple AI voice condition reported significantly lower intrinsic and extraneous cognitive load compared to those in the single AI voice condition. Furthermore, the multiple AI voice group outperformed the single AI voice group in both immediate and delayed retention, as well as in immediate far-transfer tasks and delayed near-transfer. This study empirically extends prior research on the cognitive effects of using multiple AI voices as virtual tutors in multimedia learning environments. It offers preliminary evidence that using unique voices to distinguish subtopics can benefit cognitive load and learning outcomes, with theoretical and instructional design implications for leveraging AI text-to-speech engines to simulate multiple virtual tutors for distinct instructional topics.

Keywords: AI voice; cognitive load; multimedia learning; multiple source effect; pedagogical agent; virtual tutor

1. Introduction

Learning programming algorithms is challenging for novice learners [1, 2]. Consider an introductory lesson on three closely related subtopics: if-statements, if-else statements, and nested-if statements. Although they share foundational concepts, these subtopics require learners to grasp subtle differences to fully understand distinct algorithmic concepts. Identifying nuances between different algorithms is crucial for applying, contrasting, and integrating unique codes to solve far-transfer problems [3]. However, novice learners struggle to infer and recognize these differences [2, 4], thereby hampering the construction of a correct mental model.

Multimedia learning environments are an effective tool for teaching programming algorithms [5]. They present instructional content by blending static images, animations, videos, text, sound, and speech [6]. The advent of generative AI tools allows instructors to effortlessly, rapidly, and affordably create multimedia learning environments. Virtual tutors [7–9] can effectively teach programming algorithms by presenting visual and verbal information in worked-example formats [10, 11].

1.1. Cognitive Theory of Multimedia Learning (CTML). Instructional design principles, rooted in the cognitive architecture of the human mind, govern the development of multimedia learning videos. The CTML [6, 12, 13] posits that a deep understanding of instructional material requires learners to (1) process visual and verbal information from the multimedia presentation through the visual and verbal sensory channels (eyes and ears), (2) organize this information into a coherent representation, and (3) activate prior knowledge to establish connections between newly presented information and existing knowledge schemas. The dual-channel and limited-capacity assumptions underpin CTML, suggesting that learners process graphical and auditory/text information through separate but constrained cognitive channels [12, 13].

Cognitive load theory (CLT) identifies three cognitive demands in the multimedia learning process: intrinsic, extraneous, and germane load [14–16]. Intrinsic load refers to the cognitive resources required to comprehend essential information, which is influenced by subject complexity and prior knowledge. Extraneous load arises when students expend cognitive effort on irrelevant details due to poor instructional design, while germane load involves allocating cognitive resources to learning strategies such as problem solving and metacognitive tracking.

According to the triarchic model of cognitive load [17], intrinsic, extraneous, and germane load correspond to essential, extraneous, and generative cognitive processing, respectively. Effective multimedia learning design minimizes extraneous load, optimizes intrinsic load, and promotes germane load to enhance deep understanding [6, 12, 13].

1.2. AI Voice in Multimedia Learning. According to social agency theory [18], multimedia learning presentations that incorporate favorable social cues can elicit positive social responses from learners [19–22]. This, in turn, encourages learners to invest greater cognitive effort in understanding the instructional material. The voice principle advocates for the use of human-recorded voices rather than computer-generated voices in narrating multimedia learning content [21, 23]. The mechanical quality of computer-generated voices can trigger undesirable social responses and impede cognitive processing, as learners feel less social obligation to engage with an unappealing artificial entity.

However, advancements in modern text-to-speech technology, also referred to as AI voices, have led to more natural, human-like speech synthesis. Research suggests that these enhanced AI voices can closely approximate the social and cognitive effects of human-recorded speech in multimedia learning environments [24, 25]. As a result, AI voices can be effectively employed as virtual tutors in multimedia learning environments [20].

Despite forming a social connection with an AI voice tutor [20, 24, 25], sustaining learners' attention and cognitive effort when processing the virtual agent's instructions remains challenging [26–28]. When engaging in a multimedia lesson on programming algorithms, learners may initially expend significant mental effort to follow the AI voice for the first subtopic but subsequently reduce their effort and attention for later sections. This decline in engagement may stem from the illusion that understanding one subtopic implies mastery of the entire lesson [29], a misconception that often afflicts novice learners who struggle to distinguish subtle differences between distinct algorithms [1, 2]. Additionally, learners may underestimate the cognitive effort required to comprehend information in an e-learning environment, as they are often misled by the seemingly effortless consumption of multimedia presentations that integrate graphical and verbal elements [30].

Novice learners experience cognitive overload [15] due to the overwhelming influx of unfamiliar and complex programming algorithm details [2, 3, 31]. Excessive cognitive load limits their ability to effectively organize and integrate verbal and visual information across different algorithm types within their working memory [6, 13]. This challenge arises from their lack of prior knowledge schemas, which prevents them from clearly distinguishing unique programming algorithm types and forming a comprehensive mental model [2, 3].

1.3. Multiple AI Voices as Specialist Virtual Tutors. This study is aimed at addressing these issues by employing multiple AI voices across three distinct programming algorithm subtopics. Each subtopic is assigned a unique AI voice persona, creating an array of specialist virtual tutors. This approach is grounded in two distinct yet interconnected theoretical propositions.

1.3.1. Multiple Virtual Tutors Facilitate Organization and Retrieval of Instructional Content. The first proposition suggests that assigning unique virtual tutors to specific information domains fosters instructional content organization and compartmentalization, which enhances cognitive load management and learning efficiency [32, 33]. Baylor and Ebbers [33] found that employing multiple pedagogical agents to fulfill distinct roles, such as expert and motivator, rather than having a single agent simulate both roles, led to an easier learning process and improved learning outcomes. Similarly, Wang et al. [34] demonstrated that learners who interacted with multiple virtual personas performing distinct functional roles, such as a moderator and an informational repository, maintained attention for longer durations and found it easier to access instructional content compared to those interacting with a single agent. Both studies credit these cognitive benefits to the two-agent condition's ability to compartmentalize instructional domains, thereby enhancing cognitive load efficiency and learning efficacy.

This concept aligns with Alpert et al.'s [35] intelligent tutoring system for programming lessons, which features domain specialists represented by two virtual tutors: the Language Guru and the Interface Guru. The interface design incorporates expertise cues based on social conventions, where learners typically seek specific domain knowledge from different professors [36]. Alpert et al. argued that specialist virtual tutors would be especially beneficial for novice learners, as they allow instructional content to be presented within a category-based organizational structure, facilitating information assimilation and retention [37, 38].

Additional cognitive psychology mechanisms may explain how multiple AI voices enhance information processing. Research suggests that incorporating intervals—such as pauses, delays, and spacing—into presented content can facilitate memory encoding and retrieval [39, 40]. Additionally, studies highlight the cognitive benefits of synthetic voice variability; variations in artificial voice traits, such as pauses and pitch, help organize content, thereby enhancing both recall and comprehension of auditory information [41, 42].

Another cognitive perspective worth exploring stems from speech processing research. Listeners may encode voice-specific attributes of speakers when processing spoken words, which can serve as additional cues for accessing and retrieving the linguistic and lexical content embedded in those words [43–45]. Thus, it is plausible that learners encode and utilize varied voice traits of distinct tutors—such as pitch, tone, volume, speed, accent, intonation, rhythm, and emphasis—as cues for integrating, categorizing, and retrieving domain-specific knowledge.

Given this, multiple virtual tutors specializing in different domains can provide structured cues that highlight gaps, transitions, and shifts across subtopics in a programming algorithm lesson. By offering specific markers to delineate information domains, this approach helps novice learners effectively distinguish, compare, and relate distinct concepts [46, 47]. Efficiently categorizing information streams [47] reduces cognitive load, allowing learners to allocate more mental resources toward retaining, organizing, and integrating information in working memory [12, 16, 17].

1.3.2. Multiple Virtual Tutors Enhance Attention and Elaboration of Instructional Content. The second theoretical proposition relates to the multiple source effect, first identified in persuasive messaging studies, which suggests that listeners pay closer attention and scrutinize information more thoroughly when it is presented by multiple sources rather than a single speaker [48–50]. The increased cognitive effort required to process messages from multiple sources occurs because listeners perceive each source's content as unique and independent, making it more worthy of deeper analysis [50, 51]. In contrast, when a single speaker delivers various pieces of information, listeners may expend less mental effort after processing the initial content, assuming subsequent information is similar [48–50]. Given humans' natural tendency to conserve cognitive resources [52, 53], this phenomenon extends to educational settings, where learners may adopt a less effortful cognitive stance and justify it by presuming the instructional content is repetitive, ultimately leading to an illusion of understanding [29, 30].

Studies have highlighted the challenge of virtual tutors in capturing and sustaining learners' attention [26–28]. While this may be partly due to learners' naturally short attention spans, disengagement can also stem from the monotonous instructional delivery of artificial agents lacking novelty factors. Additionally, virtual tutors' instructional prompts do not always promote deeper information processing, particularly among novices. Instead, a lack of relevant knowledge schemas and cognitive overload may lead learners to engage in shallow processing when attending to virtual tutors' messages [54].

Furthermore, while AI-generated voices may initially appear human-like and natural, their synthetic nature often remains discernible. Under complex learning demands, these artificial subtleties can reduce cognitive efficiency, impair recall performance, and lower motivation to remain engaged with the agents. This decline is attributed to a perceived lack of social cues and mental fatigue resulting from the suboptimal qualities of AI-generated speech [55, 56].

The preceding issues highlight the potential of the multiple source effect in generating cognitive benefits in multimedia learning. Individual source speakers introduce a sense of novelty, priming listeners to "gear up" when processing informational content [50]. Incorporating dynamism and novelty factors can counter inattention and disengagement in an otherwise monotonous multimedia learning environment. Thus, employing multiple AI voices as specialist virtual tutors can reduce monotony, fostering sustained attention and active processing across various learning topics over an extended period. This approach aligns with the use of unique virtual educational personas for distinct instructional roles and content, encouraging learners to process information efficiently while also revitalizing their interest and motivation to engage more attentively with the agents' message.

2. The Present Study and Hypotheses

While theoretical insights suggest that employing multiple AI voices as specialist virtual tutors for distinct subtopics could yield cognitive benefits in multimedia learning, surprisingly few empirical studies have explored this notion. Limited research provides preliminary evidence that multiple agents can reduce cognitive load and enhance learning outcomes, credited to their ability to compartmentalize information presentation [33, 34]. Alpert et al. [35] developed a two-agent intelligent tutoring system based on the premise that specialist agents can provide cues for categorizing information subdomains. However, the researchers did not conduct an experiment to compare this design against a control condition (e.g., a single-agent condition), making it difficult to establish causal effects.

Our comprehensive literature review reveals a lack of systematic studies investigating the cognitive effects of assigning specialist virtual tutors to specific subtopics within a multimedia learning environment. No follow-up research has extended the findings made over 15 years ago [32–35]. This research gap presents an opportunity to explore the potential of employing multiple AI voices as specialist virtual tutors to enhance multimedia learning.

Building on prior research suggesting that multiple virtual agents can enhance learning by reducing cognitive load through compartmentalized information presentation [33, 34], this study examines how employing multiple AI voices as specialist virtual tutors influences intrinsic, extraneous, and germane cognitive load in a multimedia learning environment. Theoretical perspectives on cognitive load suggest that segmenting instructional content among specialized tutors may reduce intrinsic load by breaking down complex information into more manageable parts while minimizing extraneous load by providing clearer structural cues for information processing [32, 35]. Furthermore, the use of multiple AI voices may enhance germane load by helping learners associate distinct tutors with specific subtopics, thereby facilitating deeper cognitive engagement and schema construction. The multiple source effect suggests that processing information from multiple distinct sources requires greater mental effort, as each source is perceived as unique, prompting deeper scrutiny and reducing the likelihood of an illusion of understanding [29, 30, 48–51]. By introducing multiple AI voices as specialist virtual tutors, instructional variation can sustain engagement, counteract monotony, and enhance cognitive effort, ultimately fostering higher germane load. Taken together, we predict cognitive load benefits of multiple AI voice tutors such that:

H1: Learners interacting with multiple AI voices as specialist virtual tutors across various subtopics will report lower intrinsic load (a), lower extraneous load (b), and higher germane load (c) than learners interacting with a single AI voice tutor.

Seminal studies suggest that multiple virtual tutors can enhance retention and knowledge transfer by compartmentalizing information and improving cognitive processing [32, 33]. By distributing instructional content across distinct AI voices, learners may benefit from clearer mental segmentation of subtopics, which can facilitate retention and transfer of knowledge, leading to better performance in both immediate and delayed posttests compared to learners interacting with a single AI voice tutor. Furthermore, the multiple source effect suggests that learners exert greater cognitive effort and engage in deeper analysis when information is presented by multiple distinct sources rather than a single speaker [48–51], which may lead to enhanced meaningful encoding and comprehension of multimedia learning materials.

In contrast, a single tutor delivering all instructional content may lead to reduced cognitive effort, disengagement, and shallow processing [26–28], resulting in an illusion of understanding rather than genuine retention and transfer of knowledge [29, 30]. Additionally, the distinct voice-based memory traces formed when listening to multiple AI tutors can enhance retrieval processes after a delay, preserving the benefits observed in immediate assessments. As a result, learners exposed to multiple voices can better recall and integrate previously learned materials, thereby outperforming those who received instruction from a single tutor on delayed posttests.

H2: Learners interacting with multiple AI voices as specialist virtual tutors across various subtopics will perform better on immediate posttest measures of retention (a), near-transfer (b), and far-transfer (c) than learners interacting with a single AI voice tutor.

Moreover, by distributing subtopics across multiple AI voices, learners can form distinct voice-specific memory traces that remain accessible over time, potentially enhancing both recall and transfer performance in delayed posttests [43–45]. In line with the categorization of instructional material [37, 38], the unique voice cues for each content area may serve as additional retrieval paths, reinforcing long-term knowledge retention and the ability to apply learned concepts in novel contexts. As a result, learners are not only more likely to recall previously learned information but also to demonstrate improved transfer performance, effectively applying knowledge to new problem-solving scenarios after a delay. We submit that this process sustains the benefits of multiple virtual tutors beyond the immediate learning phase, translating into positive effects on retention and transfer outcomes over time.

H3: Learners interacting with multiple AI voices as specialist virtual tutors across various subtopics will perform better on delayed posttest measures of retention (a), near-transfer (b), and far-transfer (c) than learners interacting with a single AI voice tutor.

3. Method

3.1. Research Design. This study employed a between-subjects experimental design, in which participants interacted with a multimedia learning video narrated by either a single AI voice tutor or three AI voice tutors. The experiment was conducted in a controlled laboratory setting. Participants completed a cognitive load survey and an immediate posttest to assess retention, near-transfer, and far-transfer performance. A delayed posttest, evaluating the same outcomes, was administered 2 weeks later.

3.2. Sampling. Using a purposive convenience sampling method, we recruited first-year business undergraduates from a private Asian university, where English is the primary language of instruction. Since these students were not specializing in information technology (IT), they likely had minimal prior knowledge of programming algorithms. Before the experiment, all participants confirmed their lack of programming knowledge through preexperiment self-assessments. Participation in the experiment was offered as an optional activity within their introductory computer applications course. Researchers briefed students on the study's purpose, emphasizing that participation was voluntary.

3.3. Multimedia Learning Environment. We developed a multimedia learning environment to teach three programming algorithm concepts: if-statement, if-else, and nested-if statements. The video includes two worked-example presentations for each algorithm, featuring narration, highlights, arrows, and animations overlaid on sample program code and a corresponding flowchart, as illustrated in Figure 1. These presentations are designed to help students interpret the output based on initial values, variables, conditions, and statements within the algorithms. Across the three algorithm subtopics, the lesson comprises six worked-example presentations, totaling about 25 min of multimedia instruction.

3.4. AI Voices as Virtual Tutors. We used Newscaster Vocalizer software to generate AI voices powered by neural networks. While these AI voices exhibit human-like inflection, they retain slight synthetic qualities yet remain clear and

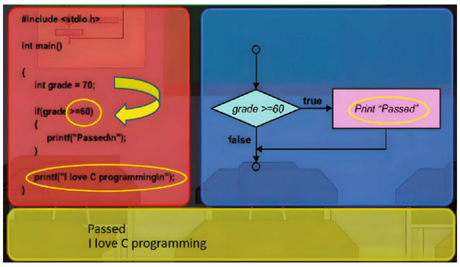

Figure 1: Multimedia learning environment.

easy to understand. To select the most suitable voices for our content, we recruited five business major undergraduates (not involved in the main experiment) to listen to sample narrations produced using various AI voices. Based on pilot test participants' consensus, we identified the top three preferred AI voices: Sonia (female), Mark (male), and Sue (female).

In the single virtual tutor multimedia learning environment, we used Sonia's voice. She introduced herself as a virtual tutor, provided an overview of the lesson, and guided learners through the three programming algorithm subtopics. In the multivirtual tutor condition, we employed three AI voices—-Sonia, Mark, and Sue—each introducing themselves as virtual tutors and providing an overview of the lesson. Sonia covered the if-statement, then introduced Mark, who led the if-else statement. Mark then introduced Sue, who took over the final subtopic, the nested-if statement.

This design was aimed at providing clear transitions between subtopics, helping to delineate and categorize the various programming algorithms. Apart from the virtual tutor introductions and distinct AI voice traits, the instructional content remained identical between the single virtual tutor and multivirtual tutor learning environments.

3.5. Dependent Measures

3.5.1. Cognitive Load Survey. We adopted Leppink et al.'s [57] 11-point Likert scale cognitive load survey to assess intrinsic load (three items), extraneous load (three items), and germane load (four items). The average scores of the relevant items represented each learner's cognitive load outcome. The internal consistency for each cognitive load scale was high, with Cronbach's alpha (α) exceeding 0.80. Table 1 presents the survey items and their corresponding Cronbach's α values.

3.5.2. Immediate Posttest. The immediate posttest was administered after learners interacted with the multimedia learning environment and completed the cognitive load survey. It assessed retention, near-transfer, and far-transfer performance. The retention test (11 marks) comprised 11 questions, each worth 1 point. Learners were instructed to recall fundamental elements and terms related to algorithmic program codes and flowcharts, such as identifying flowchart components, types of program statements, possible outcomes, and key terms in program codes. The near-transfer test (10 marks) included 10 program codes, each representing one of the three algorithms. Learners were tasked with interpreting the code output, earning 1 point per correct answer.

The far-transfer test (15 marks) consisted of three questions requiring learners to apply different algorithms to write program codes for specific objectives. For instance, in the first question, learners developed a program to calculate and display the correct BMI category based on a person's weight. The second question involved programming an air conditioning system to display and automatically adjust temperature settings. The third question required learners to design mobile game controls for combative moves based on specific button inputs. These tasks assessed learners' ability to differentiate between algorithms, select the most appropriate one, and, when necessary, integrate multiple algorithmic codes to achieve the objectives. Each far-transfer question was worth up to 5 points, based on the accuracy and quality of learners' algorithmic statements, values, and conditions, for a total of 15 marks. Sample questions for each test are provided in Appendix A.

3.5.3. Delayed Posttest. The delayed test was administered 2 weeks after the experiment and assessed retention, near-transfer, and far-transfer performance, with the number of questions and scoring system identical to the immediate posttest. While the delayed retention questions remained the same as those in the immediate posttest, the 10 delayed near-transfer questions featured different program codes. Similarly, the far-transfer questions in the delayed test were modified. The first question required learners to program an intelligent washing machine to display and automate functions. The second task involved developing a system to assess weight and determine product shipment costs. The third task required learners to classify individuals into generational demographic cohorts based on age. Sample questions for each test are provided in Appendix A.

3.6. Sampling, Participants, and Experimental Procedure. This study employed purposive convenience sampling to recruit first-year business undergraduates from a large private Asian university, where English is the primary language of instruction. A total of 42 students (19 males, 23 females), aged 18–22 years ($M = 20.5$, $SD = 1.3$), participated in the experiment. As non-IT majors, these students had minimal prior exposure to programming concepts, making them an ideal group for examining the cognitive effects of AI voice-based tutoring in an introductory learning context.

Prior to the experiment, all participants confirmed their lack of programming knowledge through preexperiment screening, ensuring a novice-level baseline for the study. Participation was voluntary and offered as an optional activity within their introductory computer applications course. To encourage participation, partial course credit was provided, though students were assured that their decision to participate or withdraw would not affect their academic standing. Since novice learners often struggle to distinguish subtle differences between algorithm types, business students were intentionally selected over computer science majors. This ensured that learners did not rely on

TABLE 1: Cognitive load survey items and Cronbach's alpha (α).

	Items	Cronbach's alpha (α)
Intrinsic	• The content of this activity was very complex. • The topics covered in this activity were very complex. • The activity content was very complex.	0.88
Extraneous	• The instructions and/or explanations during the activity were very unclear. • The instructions and/or explanations were, in terms of learning, very ineffective. • The instructions and/or explanations were full of unclear language.	0.83
Germane	• The activity really enhanced my understanding of the topics covered. • The activity really enhanced my knowledge and understanding of the topics covered. • The activity really enhanced my understanding of the content covered. • The activity really enhanced my understanding of the concepts covered.	0.86

TABLE 2: Means and standard deviations of the dependent measures across the conditions.

Dependent measures		Multiple AI voices $n = 20$ M (SD)	Single AI voice $n = 21$ M (SD)
Cognitive load	Intrinsic	3.07 (1.98)	4.67 (2.17)
	Extraneous	2.43 (1.01)	3.35 (1.22)
	Germane	7.85 (2.00)	7.32 (1.63)
Immediate posttest	Retention (max possible score = 11)	8.15 (2.30)	6.43 (2.75)
	Near-transfer (max possible score = 10)	8.05 (2.33)	7.00 (3.54)
	Far-transfer (max possible score = 15)	10.18 (3.72)	7.40 (4.85)
Delayed posttest	Retention (max possible score = 11)	3.95 (2.10)	2.64 (1.96)
	Near-transfer (max possible score = 10)	8.45 (2.61)	6.86 (3.41)
	Far-transfer (max possible score = 15)	7.18 (4.92)	5.71 (4.35)

preexisting schemas, allowing for a more precise evaluation of how AI voice–based instructional segmentation influences cognitive load and learning outcomes.

Participants were ushered into a computer laboratory and seated at desktop computers equipped with headphones. Researchers briefed them on the study's purpose, emphasized voluntary participation, and facilitated the completion of informed consent forms. Participants were then randomly assigned to either the single virtual tutor or multiple virtual tutor multimedia learning environment. The multimedia lesson lasted approximately 25 min, after which participants completed the cognitive load survey.

Following the survey, participants completed the immediate posttest in a paper-and-pencil format, assessing retention, near-transfer, and far-transfer performance. The retention test had a time limit of 3.5 min, the near-transfer test was allotted 10 min, and each far-transfer test question was given 3.5 min. Experimenters collected each answer sheet after the allocated time and distributed new questions accordingly. Including a debriefing session, the entire experiment lasted approximately 1 h.

After a 2-week interval, 41 participants (with one participant absent) returned to the same computer laboratory for the delayed posttest. The testing procedure, time limits, and protocols remained identical to the immediate posttest, covering retention, near-transfer, and far-transfer assessments. Two authors of this study scored all posttests using a rubric and answer scheme, with the scorers blinded to experimental conditions. Any minor discrepancies in scoring were resolved through consensus between the scorers.

4. Data Analyses and Results

4.1. Descriptive Data. Table 2 shows the means and standard deviations of the cognitive load measures and learning outcomes.

4.2. Data Screening and Normality Check. Prior to conducting our primary analyses, we examined the distributions of all outcome measures to assess normality and to identify any potential ceiling or floor effects. The outcomes included intrinsic load, extraneous load, and germane load; immediate retention, immediate near-transfer, and immediate far-transfer; and delayed retention, delayed near-transfer, and delayed far-transfer. For each variable, we performed Kolmogorov–Smirnov (K–S) and Shapiro–Wilk (S–W) tests separately for the single AI voice tutor (control) and multiple AI voice tutors (experimental) conditions, and we inspected *Q-Q* plots and histograms to supplement these formal tests.

The analyses revealed certain deviations from normality, potentially reflecting ceiling and/or floor effects. In the multiple AI voice tutors (experimental) condition, germane load significantly deviated from normality (S–W $p = 0.001$), exhibiting a pronounced negative skew and a median near the scale maximum, suggesting a strong ceiling effect. Immediate near-transfer scores deviated significantly from normality in both conditions ($p < 0.001$ in the single AI voice tutor condition and p between 0.001 and 0.003 in the experimental condition), indicating that many participants scored near the upper bound. Immediate far-transfer scores were also nonnormal in the experimental condition (K–S $p = 0.005$, S–W $p = 0.014$) despite being normally distributed in the single AI voice tutor condition. Among the delayed measures, delayed retention in the single AI voice tutor condition was nonnormal (S–W $p = 0.020$). Delayed near-transfer scores significantly deviated from normality in both conditions, with a strong negative skew in the experimental group (skewness = −2.298), while delayed far-transfer scores remained normally distributed in both groups.

These ceiling and floor effects may be attributed to inherent task and measurement scale characteristics. For instance, the near-transfer tasks appeared relatively easy for many participants, resulting in scores clustering at the upper limit, while the delayed retention task proved more challenging for a subset of participants, leading to scores near the lower bound. Despite statistical evidence of nonnormality, visual inspection of *Q-Q* plots and histograms did not reveal extreme outliers. Given our balanced sample sizes (approximately 20–21 participants per condition) and the robustness of ANOVA and MANOVA to moderate departures from normality, we proceeded with parametric analyses. However, we acknowledge that the observed ceiling and floor effects may have influenced variability and effect size estimates, and we interpreted our results accordingly in light of these distributional characteristics.

4.3. Hypothesis Testing. This study predicts that multiple AI voices, acting as specialist virtual tutors across various subtopics, will result in lower intrinsic load, lower extraneous load, and higher germane load compared to a single AI voice tutor. Additionally, it is hypothesized that multiple AI voices will lead to better immediate and delayed learning performance than a single AI voice tutor.

To test these predictions, we conducted a series of MANOVAs to compare cognitive load, immediate posttests, and delayed posttests across conditions. This approach accounts for the interrelated nature of the dependent variables and the need for a comprehensive statistical method. Cognitive load (*intrinsic*, *extraneous*, and *germane*) is conceptually linked, as are the immediate posttest components (*retention*, *near-transfer*, and *far-transfer*) and the delayed posttest components measuring the same outcomes over time. Given these interdependencies, MANOVA was chosen to analyze multiple outcomes simultaneously while minimizing the risk of *Type I* errors from conducting separate ANOVAs.

For significant MANOVA results, we conducted post hoc analyses using Storey's *False Discovery Rate* (*FDR*) procedure in *R* [58] to control for multiple comparisons while preserving statistical power. This approach has become increasingly prevalent in educational research [59–61], particularly when sample sizes are limited. It serves as an alternative to *family-wise error rate* (*FWER*) corrections, such as the *Bonferroni method*, which, while effective in controlling *Type I* errors, can be overly conservative, increasing the risk of *Type II* errors and potentially obscuring meaningful effects.

Storey's *FDR* method first computes *one-tailedp values* for each hypothesis, reflecting the directional nature of the tests, and then estimates the proportion of true null hypotheses by analyzing the *p value* distribution and selecting a threshold (λ) to distinguish signal from noise [58]. Using this estimate, *q-values* are calculated to determine the minimum *FDR* at which a hypothesis is deemed significant. Unlike *p values*, which assess individual tests, *q -values* adjust for multiple comparisons, ensuring a data-driven significance threshold that preserves statistical power [62].

In this study, Storey's *FDR* was applied via the *q-value* package in *R*, with *q-values* computed based on the observed *p value* range as the λ parameter. A result was deemed significant if both the *one-tailedp value* and the corresponding *q value* were below 0.05, balancing *Type I* and *Type II* error control [63]. This approach effectively adjusts for multiple comparisons while avoiding the overly conservative nature of *FWER* corrections [62, 63], making it particularly advantageous in studies where preserving statistical power is critical.

4.3.1. Cognitive Load Ratings. A one-way MANOVA was conducted to examine the effect of AI voice condition (multiple AI voices as specialist virtual tutors vs. a single AI voice tutor) on cognitive load ratings (*intrinsic*, *extraneous*, and *germane cognitive load*). Box's test of equality of covariance matrices was not significant ($M = 5.57$, $F(6, 11288.73) = 0.85$, $p = 0.52$), and Levene's test for equality of error variances was also not significant for all dependent variables ($F \leq 0.21$, $p \geq 0.65$), indicating that the assumptions of homogeneity of covariance matrices and error variances were met. A statistically significant effect of AI voice condition on cognitive load was detected ($Wilk's\Lambda = 0.790$, $F(3, 38) = 3.37$, $p < 0.05$, $\eta^2 = 0.21$).

Following this significant multivariate result, separate univariate ANOVAs were conducted for *intrinsic*, *extraneous*, and *germane cognitive load*, yielding one-tailed p values of 0.0055, 0.009, and 0.1595, respectively. The effect sizes (*Cohen'sd*) indicated a moderate-to-large effect on intrinsic load ($d = 0.78$) and a large effect on extraneous load ($d = 0.85$), suggesting that multiple AI voices significantly reduced cognitive load for these variables. However, the effect on germane load was small and nonsignificant ($d = 0.32$).

To account for multiple comparisons, Storey's *FDR* method was applied. When compared to an α level of 0.05, the results confirmed that the effects on extraneous and intrinsic load remained statistically significant ($p < 0.05$), whereas the effect on germane load was not significant ($p > 0.05$). Taken together, these findings partially support H1, as multiple AI voices acting as specialist virtual tutors across subtopics resulted in lower intrinsic and extraneous load compared to a single AI voice tutor. However, germane load was not enhanced as expected.

4.3.2. Immediate Posttest Scores. A one-way MANOVA was conducted to examine the effect of AI voice condition (multiple AI voices as specialist virtual tutors vs. a single AI voice tutor) on immediate posttest scores, measured by retention, near-transfer, and far-transfer performance. Box's test of equality of covariance matrices was not significant ($M = 10.49$, $F(6, 11288.74) = 1.61$, $p = 0.14$), indicating that the assumption of homogeneity of covariance matrices was met. However, Levene's test of equality of error variances indicated a violation for near-transfer ($F(1, 40) = 12.40$, $p = 0.00$), while variances for retention ($p = 0.22$) and far-transfer ($p = 0.12$) were not significantly different across groups. Despite this partial violation, the overall MANOVA was statistically significant ($Wilk's\Lambda = 0.81$, $F(3, 38) = 2.99$, $p < 0.05$, $\eta^2 = 0.19$).

Following this significant multivariate result, separate univariate ANOVAs were conducted for retention, near-transfer, and far-transfer, yielding one-tailed p values of 0.0135, 0.1655, and 0.024, respectively. The effect sizes (*Cohen'sd*) indicated a moderate-to-large effect on retention ($d = 0.73$) and far-transfer ($d = 0.64$), suggesting that multiple AI voices significantly enhanced these learning outcomes. However, the effect on near-transfer was small and nonsignificant ($d = 0.31$).

To account for multiple comparisons, Storey's *FDR* method was applied. When compared to an α level of 0.05, the results confirmed that the effects on retention and far-transfer remained statistically significant ($p < 0.05$), whereas the effect on near-transfer was not significant ($p > 0.05$). Thus, these analyses provide partial support for H2, as multiple AI voices acting as specialist virtual tutors across subtopics resulted in higher immediate retention and far-transfer scores but did not enhance near-transfer scores compared to a single AI voice tutor.

4.3.3. Delayed Posttest Scores. A one-way MANOVA was conducted to examine the effect of AI voice condition (multiple AI voices as specialist virtual tutors vs. a single AI voice tutor) on delayed posttest scores, measured by delayed recall, delayed near-transfer, and delayed far-transfer performance. Box's test of equality of covariance matrices was not significant ($M = 8.14$, $F(6, 10942.89) = 1.24$, $p = 0.28$), indicating that the assumption of homogeneity of covariance matrices was met. However, Levene's test of equality of error variances revealed a violation for delayed near-transfer ($F(1, 39) = 4.65$, $p = 0.04$), while the variances for delayed recall ($p = 0.42$) and delayed far-transfer ($p = 0.34$) were not significantly different across groups. Despite this partial violation, the overall MANOVA based on $Wilk's\Lambda$ was statistically significant ($\Lambda = 0.87$, $F(3, 37) = 1.91$, $p < 0.05$, $\eta^2 = 0.13$).

Following this significant multivariate result, separate univariate ANOVAs were conducted for delayed recall, delayed near-transfer, and delayed far-transfer, yielding one-tailed p values of 0.02, 0.05, and 0.15, respectively. The effect sizes (*Cohen'sd*) indicated a moderate-to-large effect on delayed recall ($d = 0.66$) and a moderate effect on delayed near-transfer ($d = 0.54$), suggesting that multiple AI voices significantly enhanced these learning outcomes. However, the effect on delayed far-transfer was small and nonsignificant ($d = 0.32$).

To account for multiple comparisons, Storey's *FDR* method was applied. When compared to an α level of 0.05, the results confirmed that the effects on delayed retention and delayed near-transfer remained statistically significant ($p < 0.05$), whereas the effect on delayed far-transfer was not significant ($p > 0.05$). Hence, the data partially supported H3, indicating that multiple AI voices acting as specialist virtual tutors across various subtopics enhanced delayed retention and near-transfer scores but did not improve delayed far-transfer scores compared to a single AI voice tutor.

4.3.4. Hypothesis Result Summary. Table 3 summarizes the hypothesis testing results, comparing the effects of multiple AI voices versus a single AI voice tutor on cognitive load components, including intrinsic load, extraneous load, and germane load, as well as learning outcomes, including immediate and delayed retention, near-transfer, and far-transfer performance. Effect size interpretations are provided to indicate the magnitude of the observed differences across these measures.

5. Discussion

This study explored the cognitive benefits of assigning multiple AI voices as specialist virtual tutors to specific programming algorithms to enhance multimedia learning outcomes. Inspired by prior research, which advocates for multiple virtual tutors to represent distinct roles and knowledge domains rather than relying on a single tutor [32–35], this study examined whether AI voice–driven instructional segmentation improves learning efficiency. Our findings suggest that multiple AI voices tailored to specific programming algorithm subtopics can optimize cognitive load and enhance learning outcomes in a multimedia learning environment.

Our data demonstrated that multiple AI voices as specialist virtual tutors had a positive impact on cognitive load. Learners who engaged with the multiple virtual tutor interface reported significantly lower perceived difficulty related to subject complexity and instructional presentation, with medium and large effect sizes, respectively. The cognitive load reduction observed in this study aligns with prior findings, where learners reported an easier learning experience and improved information accessibility when using a two-agent interface compared to a single-agent design [33, 34]. These findings suggest that implementing multiple AI voices can effectively manage cognitive load across multiple subtopics, a critical factor in successful multimedia learning.

Beyond cognitive load reductions, the cognitive benefits of employing multiple AI voices as specialist virtual tutors extend to learning performance. In the immediate posttest, learners in the multiple virtual tutor condition demonstrated higher retention scores than those in the single virtual tutor condition, with a moderate to large effect. This superior retention persisted in the 2-week delayed posttest, despite

TABLE 3: Summary of hypothesis testing results.

Hypothesis	Prediction	Results	Effect size
H1 Multiple AI voices > single AI voice	Intrinsic load ↓	Supported	Moderate to large effect
	Extraneous load ↓	Supported	Large effect
	Germane load ↑	Not supported	Small effect
H2 Multiple AI voices > single AI voice	Immediate retention ↑	Supported	Moderate-to-large effect
	Immediate near-transfer ↑	Not supported	Small effect
	Immediate far-transfer ↑	Supported	Moderate-to-large effect
H3 Multiple AI voices > single AI voice	Delayed retention ↑	Supported	Moderate-to-large effect
	Delayed near-transfer ↑	Supported	Moderate effect
	Delayed far-transfer ↑	Not supported	Small effect

Note: The arrows (↑, ↓) indicate the predicted direction of change in the measured variable. "↑" denotes a predicted increase, whereas "↓" denotes a predicted decrease.

the expected memory decay over time. Notably, learners in the multiple virtual tutor environment continued to outperform their single virtual tutor counterparts, with a moderate effect for both delayed retention and delayed near-transfer. These findings collectively highlight the robust impact of multiple AI voices on learning performance, reinforcing prior research that found higher recall performance with a two-agent interface compared to a single-agent interface [33].

In detail, both immediate and delayed near-transfer scores followed different patterns across conditions. While immediate near-transfer scores did not significantly differ between the multiple virtual tutor and single virtual tutor conditions, delayed near-transfer scores revealed a significant advantage for learners in the multiple AI voice condition. Learners in both conditions initially achieved relatively high scores on the immediate near-transfer test, suggesting that the questions may have been too easy, leading to a potential ceiling effect. However, in the delayed test, learners who engaged with multiple AI voices demonstrated superior near-transfer performance, indicating a long-term learning benefit. Conversely, the far-transfer test was more challenging, requiring learners to identify, contrast, compare, correlate, and integrate different algorithms to complete tasks dissimilar to those presented during the learning phase. Learners who engaged with multiple AI voice tutors outperformed those in the single AI voice condition on the immediate far-transfer test, showing a medium effect size, though this effect did not persist in the delayed posttest.

While the cognitive benefits associated with cognitive load and learning outcomes derived from multiple AI voices align with prior findings [33, 34], it is important to highlight the unique aspects of this study. Baylor and Ebbers [33] and Wang et al. [34] employed multiple embodied virtual agents to simulate differential roles, such as domain expertise information, motivational support, or moderation prompts. In contrast, our multivirtual tutor design focused on delivering specific domain information segmented by subtopics rather than assigning different instructional roles. Another distinct feature of this study is the use of Leppink et al.'s [57] scale, which allowed us to decompose cognitive load effects into intrinsic, extraneous, and germane loads, thereby providing greater clarity on how multiple AI voices influence different types of cognitive load in multimedia learning.

Furthermore, by dividing the posttest into retention, near-transfer, and far-transfer assessments, we gained a more detailed understanding of how multiple AI voices influenced learning outcomes. The inclusion of a far-transfer test allowed for a more sensitive evaluation of learners' depth of understanding, while the delayed posttest enabled us to assess long-term retention effects. Synthesizing these posttest data highlights three key learning benefits of multiple AI voices: (1) enhanced retention, linked to improved encoding, storage, and retrieval of algorithmic concepts; (2) improved near-transfer performance, particularly in the delayed posttest, suggesting that multiple AI voices facilitated sustained application of learned concepts over time; and (3) superior far-transfer performance, observed in the immediate posttest, where learners were better able to adapt and apply different algorithms to novel tasks. Taken together, this study builds on prior research by providing empirical evidence that utilizing multiple AI voices as specialized virtual tutors across subtopics offers cognitive benefits by reducing difficulty ratings related to subject complexity (intrinsic load) and instructional presentation (extraneous load) while enhancing retention, delayed near-transfer, and immediate far-transfer performance.

From a theoretical perspective, the cognitive load reduction and enhanced learning performance facilitated by multiple AI voices support the proposition that assigning virtual tutors as specialists across subtopics promotes information organization and compartmentalization [32–35]. By orchestrating three AI voices as distinct virtual tutor personas, each introducing themselves, teaching a specific algorithm, and passing the next algorithm topic to the next tutor, learners are guided to infer gaps, transitions, and contrasts across subtopics. This structure promotes memorization of information chunks and assimilates new knowledge [46, 47].

Additionally, drawing from speech processing research [44, 45], it can be theorized that the unique voice qualities of specialist virtual tutors are encoded alongside instructional material, serving as cues for content categorization and facilitating information retrieval from long-term memory. In line with CLT [16, 17], these factors may assist in managing essential and generative cognitive processing in multimedia learning, optimizing cognitive resources for deeper processing of instructional material [13, 47]. This, in turn, enhances retention and far-transfer performance while leading to lower intrinsic and extraneous load ratings.

Another theoretical perspective that may explain the cognitive benefits of multiple AI voices in this study is the multiple-source effect [48–50]. This theory aligns with the superior retention and far-transfer scores observed in learners who interacted with multiple AI voices. Learners may perceive instructional content delivered by distinct virtual tutors as unique and worthy of closer scrutiny, prompting greater attention and deeper elaboration on the three algorithms. This process discourages the illusion of understanding, encouraging learners to deliberate on each algorithm individually [29, 30].

The increased mental effort in elaborating on each algorithm likely enabled learners in the multiple virtual tutor condition to excel in identifying, contrasting, comparing, correlating, and integrating different algorithms in the far-transfer task. Additionally, it is plausible that learners who listened to a single AI voice experienced dissonance and fatigue due to the perceived artificiality of the voice or the monotony of the presentation—both of which could negatively impact cognitive load and learning outcomes [55, 56]. In contrast, introducing novel personas through multiple AI voices may dispel monotony, sustain learner interest, and revitalize motivation, ultimately enhancing cognitive load management and improving learning outcomes across subtopics.

6. Practical Implications for Education

Considering the cognitive advantages of multiple AI voices, it is essential to shift the discussion toward practical considerations for leveraging AI voices to enhance multimedia learning. Modern text-to-speech vocalizers offer a diverse array of AI voice personas, enabling instructional designers to create engaging multimedia lessons. Notably, the multiple virtual tutor interface is particularly useful for multimedia lessons on highly concentric subtopics, such as programming algorithms. This recommendation is based on the principle that multiple AI voices as specialist virtual tutors help learners distinguish nuanced differences among subtopics while fostering sustained attention and active engagement with instructional materials over time.

Another practical consideration involves the selection of AI voices for multimedia learning. Choosing AI voices that align with learners' expectations is crucial for evoking favorable social, emotional, and motivational responses. This includes various voice traits such as pitch, tone, volume, speed, accent, intonation, rhythm, and emphasis, which collectively shape a tutor persona that enhances the learning experience. Additionally, different demographics may respond differently to various voice traits and personas, making it essential to tailor AI voice choices to match the preferences and characteristics of target learners.

Instructional designers should also ensure dissimilarity among AI voice profiles to create distinct virtual personas. This diversity allows each virtual tutor to provide unique cues that aid learners in categorizing information, encoding content, and retrieving knowledge from memory. By optimizing cognitive load processing, this approach enhances learning outcomes and promotes more effective multimedia instruction.

7. Limitations and Future Outlook

This study, while providing insights into the cognitive benefits of employing multiple AI voices as specialist virtual tutors, is not without limitations. One key constraint is its specific subject domain of introductory programming algorithms, which may not directly translate to other subjects or educational levels. Additionally, the sample size was relatively small, comprising first-year business undergraduates from a private Asian university where English is the primary language of instruction. This narrow demographic focus limits the generalizability of the findings to other cultures, age groups, or academic disciplines.

Future research is needed to address these limitations and further explore the long-term effects of learning with multiple AI voices beyond a single engagement session. This could provide insights into whether habituation effects might influence the instructional strategy's impact. While this study offers empirical support for cognitive psychological propositions regarding the interplay of multiple AI voices and multimedia learning processes, our experimental design does not fully unravel the exact mechanisms contributing to cognitive load reduction and learning outcome improvements. Future studies should investigate alternative learning theories, particularly those focusing on social and affective constructs related to AI voice variability. Additionally, exploring the effectiveness of this approach across different disciplines and more diverse learner populations would be valuable.

Despite its limitations, this study contributes to the existing body of research on multimedia learning by examining the use of multiple AI voices as specialist virtual tutors. The findings suggest that this approach reduces cognitive load and enhances learning outcomes, encouraging further exploration of multimedia learning strategies that leverage technological advancements in generative AI tools, such as AI voice synthesizers.

Appendix A: Sample Posttest Questions for Retention, Near-Transfer, and Far-Transfer

1. Retention question

Label the following components in the program:

2. Immediate near-transfer test

Based on the following source code, determine and write the expected output.

3. Immediate far-transfer test

Imagine you are a game developer working on a mobile fighting game featuring a hero fighter. The game includes

```
int main ()
{
        int temperature = 25;
        if (temperature >= 30)
        {
           printf("It's a hot day!");
        }
        printf("Stay hydrated.");
}
```

Code 1

```
#include <stdio.h>
int main()
{
        int heartbeat = 85;
        if (heartbeat > 100)
        {
           printf("Your heart rate is above normal.\n");
           printf("Please consult a doctor.\n");
        }
        else if (heartbeat < 60)
        {
           printf("Your heart rate is below normal.\n");
           printf("Please consult a doctor.\n");
        }
        else
        {
           printf("Your heart rate is within the normal
range.\n");
        }
        printf("Thank you.\n");
        return 0;
}
```

Code 2

three action buttons: "X," "Y," and "Z," each triggering a special attack when pressed multiple times.

In this game, if the "X" button is pressed more than two times, the screen will display "Attack Damage: 120" followed by "DAMAGE." If the "Y" button is pressed more than three times, the screen will display "Attack Damage: 300" followed by "DAMAGE." Similarly, if the "Z" button is pressed more than five times, the screen will display "Attack Damage: 1000" followed by "DAMAGE."

Your task is to write a simple C program that simulates the hero's attack system. The program should prompt the player to enter the number of times they pressed each button (X, Y, or Z) and then determine and display the appropriate attack damage based on the given conditions.

4. Delayed near-transfer

Based on the following source code, determine and write the expected output.

5. Delayed far-transfer

```
#include <stdio.h>
int main()
{
        int blood_pressure = 110;
        if (blood_pressure >= 120)
        {
           printf("Blood pressure is normal.\n");
           printf("Please see the nurse for a routine check-
up.\n");
        }
        else if (blood_pressure < 90)
        {
           printf("Blood pressure is too low.\n");
           printf("Please seek medical attention.\n");
        }
        else
        {
           printf("Blood pressure is slightly low.\n");
           printf("Please monitor your health.\n");
        }
        printf("Thank you.\n");
        return 0;
}
```

Code 3

A modern IoT-enabled smart washing machine uses sensors to detect laundry weight and display appropriate messages. By default, the screen shows "Ready Standby" followed by "Please load clothes." If the detected weight reaches or exceeds 30 kg, the display changes to "Load nearing full capacity" followed by "Please remove excess clothes if needed."

As a programmer, write a simple C program that prompts the user to enter the laundry weight, checks if it is 30 kg or more, and displays the correct message. Use conditional statements to ensure accurate output.

Data Availability Statement

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

Ethics Statement

This study was conducted in accordance with the ethical principles outlined in the Helsinki Declaration, ensuring the protection of participants' rights, dignity, and privacy. In line with national regulations and institutional policies, formal ethical approval was not required, as the study did not involve the collection of personally identifiable or sensitive data. All participants provided informed consent electronically before participation. They were fully briefed on the study's objectives, procedures, and their rights, including their right to withdraw at any time without penalty or consequences. Additionally, the study complied with the Personal Data Protection Act (PDPA), ensuring that all collected data was handled securely, used exclusively for research purposes, and protected against unauthorized access or misuse through strict confidentiality measures.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

The authors acknowledge the support of the Research Management Centre of Multimedia University through the IRFUND grant (ID: MMUI/240123) awarded to the first author.

References

[1] C. S. Cheah, "Factors Contributing to the Difficulties in Teaching and Learning of Computer Programming: A Literature Review," *Contemporary Educational Technology* 12, no. 2 (2020): ep272, https://doi.org/10.30935/cedtech/8247.

[2] Y. Qian and J. Lehman, "Students' Misconceptions and Other Difficulties in Introductory Programming: A Literature Review," *ACM Transactions on Computing Education (TOCE)* 18, no. 1 (2017): 1–24.

[3] R. Weeda, S. Smetsers, and E. Barendsen, "Unraveling Novices' Code Composition Difficulties," *Computer Science Education* (pp. 1–28.

[4] P. Kather and J. Vahrenhold, "Is Algorithm Comprehension Different From Program Comprehension?," in *Proceedings of 2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC)* (IEEE, 2021), 455–466, https://doi.org/10.1109/ICPC52881.2021.00053.

[5] M. D. Abdulrahaman, N. Faruk, A. A. Oloyede, et al., "Multimedia Tools in the Teaching and Learning Processes: A Systematic Review," *Heliyon* 6, no. 11 (2020): e05312, https://doi.org/10.1016/j.heliyon.2020.e05312.

[6] R. C. Clark and R. E. Mayer, *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* (John Wiley & Sons, 5th ed. edition, 2023).

[7] T. W. Liew, S. M. Tan, and C. Jayothisa, "The Effects of Peer-Like and Expert-Like Pedagogical Agents on Learners' Agent Perceptions, Task-Related Attitudes, and Learning Achievement," *Journal of Educational Technology & Society* 16, no. 4 (2013): 275–286.

[8] T. W. Liew, N. A. M. Zin, and N. Sahari, "Exploring the Affective, Motivational, and Cognitive Effects of Pedagogical Agent Enthusiasm in a Multimedia Learning Environment," *Human-Centric Computing and Information Sciences* 7, no. 1 (2017): 1–21, https://doi.org/10.1186/s13673-017-0089-2.

[9] A. M. Sinatra, K. A. Pollard, B. T. Files, A. H. Oiknine, M. Ericson, and P. Khooshabeh, "Social Fidelity in Virtual Agents: Impacts on Presence and Learning," *Computers in Human Behavior* 114 (2021): 106562.

[10] S. Bichler, M. Schwaighofer, M. Stadler, M. Bühner, S. Greiff, and F. Fischer, "How Working Memory Capacity and Shifting Matter for Learning With Worked Examples—A Replication Study," *Journal of Educational Psychology* 112, no. 7 (2020): 1320–1335, https://doi.org/10.1037/edu0000423.

[11] C. Wu, J. Deboer, J. F. Rhoads, and E. Berger, "Use of Worked-Example Videos to Support Problem-Solving: An Analysis of Student Behavior," *Computer Applications in Engineering Education* 30, no. 1 (2022): 195–221.

[12] R. E. Mayer and R. Moreno, "A Cognitive Theory of Multimedia Learning: Implications for Design Principles," *Journal of Educational Psychology* 91, no. 2 (1998): 358–368.

[13] R. E. Mayer and R. Moreno, "Nine Ways to Reduce Cognitive Load in Multimedia Learning," *Educational Psychologist* 38, no. 1 (2003): 43–52.

[14] F. Krieglstein, M. Beege, G. D. Rey, P. Ginns, M. Krell, and S. Schneider, "A Systematic Meta-Analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research," *Educational Psychology Review* 34, no. 4 (2022): 2485–2541, https://doi.org/10.1007/s10648-022-09683-4.

[15] D. Mutlu-Bayraktar, V. Cosgun, and T. Altan, "Cognitive Load in Multimedia Learning Environments: A Systematic Review," *Computers & Education* 141 (2019).

[16] J. Sweller, J. J. G. van Merriënboer, and F. G. Paas, "Cognitive Architecture and Instructional Design," *Educational Psychology Review* 10 (1998): 251–296.

[17] S. Kalyuga, "Cognitive Load Theory: How Many Types of Load Does It Really Need?," *Educational Psychology Review* 23, no. 1 (2011): 1–19, https://doi.org/10.1007/s10648-010-9150-7.

[18] R. E. Mayer, "Principles Based on Social Cues in Multimedia Learning: Personalization, Voice, Image, and Embodiment Principles," in *The Cambridge Handbook of Multimedia Learning* 16, Cambridge University Press, 2014), 345–370.

[19] M. Beege and S. Schneider, "Emotional Design of Pedagogical Agents: The Influence of Enthusiasm and Model-Observer Similarity," *Educational Technology Research and Development* 71, no. 3 (2023): 859–880, https://doi.org/10.1007/s11423-023-10213-4.

[20] T. W. Liew, S. M. Tan, W. M. Pang, M. T. I. Khan, and S. N. Kew, "I Am Alexa, Your Virtual Tutor!: The Effects of Amazon Alexa's Text-to-Speech Voice Enthusiasm in a Multimedia Learning Environment," *Education and Information Technologies* 28, no. 2 (2023): 1455–1489, https://doi.org/10.1007/s10639-022-11255-6.

[21] T. W. Liew, S. M. Tan, T. M. Tan, and S. N. Kew, "Does Speaker's Voice Enthusiasm Affect Social Cue, Cognitive Load, and Transfer in Multimedia Learning?," *Information and Learning Sciences* 121 (2020): 117–135, https://doi.org/10.1108/ILS-11-2019-0124.

[22] S. Schneider, M. Beege, S. Nebel, L. Schnaubert, and G. D. Rey, "The Cognitive-Affective-Social Theory of Learning in Digital Environments (CASTLE)," *Educational Psychology Review* 34, no. 1 (2022): 1–38.

[23] R. K. Atkinson, R. E. Mayer, and M. M. Merrill, "Fostering Social Agency in Multimedia Learning: Examining the Impact of an Animated Agent's Voice," *Contemporary Educational Psychology* 30, no. 1 (2005): 117–139, https://doi.org/10.1016/j.cedpsych.2004.07.001.

[24] S. D. Craig and N. L. Schroeder, "Reconsidering the Voice Effect When Learning From a Virtual Human," *Computers & Education* 114 (2017): 193–205, https://doi.org/10.1016/j.compedu.2017.07.013.

[25] S. D. Craig and N. L. Schroeder, "Text-to-Speech Software and Learning: Investigating the Relevancy of the Voice Effect," *Journal of Educational Computing Research* 57, no. 6 (2019): 1534–1548, https://doi.org/10.1177/0735633118806282.

[26] Y. Hayashi, "On Pedagogical Effects of Learner-Support Agents in Collaborative Interaction," in *ITS'12: Proceedings of the 11th international conference on Intelligent Tutoring Systems* 11, Springer, 2012), 22–32, https://doi.org/10.1007/978-3-642-30950-2_3.

[27] Y. Hayashi, "Togetherness: Multiple Pedagogical Conversational Agents as Companions in Collaborative Learning," in *Intelligent Tutoring Systems: 12th International Conference, ITS 2014* 12, Springer, 2014), 114–123, https://doi.org/10.1007/978-3-319-07221-0_15.

[28] Y. Hayashi, "Multiple Pedagogical Conversational Agents to Support Learner-Learner Collaborative Learning: Effects of Splitting Suggestion Types," *Cognitive Systems Research* 54 (2019): 246–257, https://doi.org/10.1016/j.cogsys.2018.04.005.

[29] M. M. Avhustiuk, I. D. Pasichnyk, and R. V. Kalamazh, "The Illusion of Knowing in Metacognitive Monitoring: Effects of the Type of Information and of Personal, Cognitive, Metacognitive, and Individual Psychological Characteristics," *Europe's Journal of Psychology* 14, no. 2 (2018): 317–333, https://doi.org/10.5964/ejop.v14i2.1526.

[30] E. S. Paik and G. Schraw, "Learning With Animation and Illusions of Understanding," *Journal of Educational Psychology* 105, no. 2 (2013): 278.

[31] Ü. Çakiroğlu, S. S. Suiçmez, Y. B. Kurtoğlu, A. Sari, S. Yildiz, and M. Öztürk, "Exploring Perceived Cognitive Load in Learning Programming via Scratch," *Research in Learning Technology* 26 (2018): https://doi.org/10.25304/rlt.v26.1888.

[32] A. L. Baylor, "Promoting Motivation With Virtual Agents and Avatars: Role of Visual Presence and Appearance," *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, no. 1535 (2009): 3559–3565, https://doi.org/10.1098/rstb.2009.0141.

[33] A. L. Baylor and S. Ebbers, "Evidence That Multiple Agents Facilitate Greater Learning," in *Artificial Intelligence in Education: Shaping the Future of Learning Through Intelligent Technologies* (IOS Press, 2003), 377–379.

[34] H. Wang, M. Chignell, and M. Ishizuka, "Are Two Talking Heads Better Than One? When Should We Use More Than One Agent in e-Learning," in *Proceedings of the 11th International Conference on Intelligent User Interfaces* (Association for Computing Machinery, 2006), 366–368, https://doi.org/10.1145/1111449.1111539.

[35] S. R. Alpert, M. K. Singley, and J. M. Carroll, "Multiple Multimodal Mentors: Delivering Computer-Based Instruction via Specialized Anthropomorphic Advisors," *Behaviour & Information Technology* 14, no. 2 (1995): 69–79, https://doi.org/10.1080/01449299508914629.

[36] T. W. Liew and S. M. Tan, "Social Cues and Implications for Designing Expert and Competent Artificial Agents: A Systematic Review," *Telematics and Informatics* 65 (2021): 101721, https://doi.org/10.1016/j.tele.2021.101721.

[37] G. H. Bower, M. C. Clark, A. M. Lesgold, and D. Winzenz, "Hierarchical Retrieval Schemes in Recall of Categorized Word Lists," *Journal of Verbal Learning and Verbal Behavior* 8, no. 3 (1969): 323–343, https://doi.org/10.1016/S0022-5371(69)80124-6.

[38] P. A. Ornstein and T. Trabasso, "To Organize Is to Remember: The Effects of Instructions to Organize and to Recall," *Journal of Experimental Psychology* 103, no. 5 (1974): 1014.

[39] W. A. Johnston and C. N. Uhl, "The Contributions of Encoding Effort and Variability to the Spacing Effect on Free Recall," *Journal of Experimental Psychology: Human Learning and Memory* 2, no. 2 (1976): 153.

[40] S. A. Madigan, "Intraserial Repetition and Coding Processes in Free Recall," *Journal of Verbal Learning and Verbal Behavior* 8, no. 6 (1969): 828–835, https://doi.org/10.1016/S0022-5371(69)80050-2.

[41] C. A. Henderson and Y. He, "Screen Reader Voices: Effects of Pauses and Voice Changes on Comprehension," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (SAGE Publications, 2022), 1839–1843, https://doi.org/10.1177/1071181322661291.

[42] R. F. Lorch Jr., H. T. Chen, and J. Lemarié, "Communicating Headings and Preview Sentences in Text and Speech," *Journal of Experimental Psychology: Applied* 18, no. 3 (2012): 265–276, https://doi.org/10.1037/a0029547.

[43] W. D. Goh, "Talker Variability and Recognition Memory: Instance-Specific and Voice-Specific Effects," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31, no. 1 (2005): 40–50, https://doi.org/10.1037/0278-7393.31.1.40.

[44] J. W. Mullennix, D. B. Pisoni, and C. S. Martin, "Some Effects of Talker Variability on Spoken Word Recognition," *The Journal of the Acoustical Society of America* 85, no. 1 (1989): 365–378.

[45] M. S. Sommers and J. Barcroft, "An Integrated Account of the Effects of Acoustic Variability in First Language and Second Language: Evidence From Amplitude, Fundamental Frequency, and Speaking Rate Variability," *Applied PsychoLinguistics* 28, no. 2 (2007): 231–249.

[46] D. Alpizar, O. O. Adesope, and R. M. Wong, "A Meta-Analysis of Signaling Principle in Multimedia Learning Environments," *Educational Technology Research and Development* 68, no. 5 (2020): 2095–2119, https://doi.org/10.1007/s11423-020-09748-7.

[47] P. D. Mautone and R. E. Mayer, "Signaling as a Cognitive Guide in Multimedia Learning," *Journal of Educational Psychology* 93, no. 2 (2001): 377–389, https://doi.org/10.1037/0022-0663.93.2.377.

[48] S. G. Harkins and R. E. Petty, "Effects of Source Magnification of Cognitive Effort on Attitudes: An Information-Processing View," *Journal of Personality and Social Psychology* 40, no. 3 (1981): 401–413, https://doi.org/10.1037/0022-3514.40.3.401.

[49] S. G. Harkins and R. E. Petty, "The Multiple Source Effect in Persuasion: The Effects of Distraction," *Personality and Social Psychology Bulletin* 7, no. 4 (1981): 627–635, https://doi.org/10.1177/014616728174005.

[50] S. G. Harkins and R. E. Petty, "Information Utility and the Multiple Source Effect," *Journal of Personality and Social Psychology* 52, no. 2 (1987): 260–268, https://doi.org/10.1037/0022-3514.52.2.260.

[51] R. E. Petty and J. T. Cacioppo, "The Effects of Involvement on Responses to Argument Quantity and Quality: Central and Peripheral Routes to Persuasion," *Journal of Personality and Social Psychology* 46, no. 1 (1984): 69.

[52] A. Shenhav, S. Musslick, F. Lieder, et al., "Toward a Rational and Mechanistic Account of Mental Effort," *Annual Review of Neuroscience* 40 (2017): 99–124.

[53] J. Shepherd, "Conscious Cognitive Effort in Cognitive Control," *Wiley Interdisciplinary Reviews: Cognitive Science* 14, no. 2 (2023): 1629.

[54] G. Trevors, M. Duffy, and R. Azevedo, "Note-Taking Within Meta Tutor: Interactions Between an Intelligent Tutoring System and Prior Knowledge on Note-Taking and Learning," *Educational Technology Research and Development* 62 (2014): 507–528.

[55] T. W. Morris and H. T. M. Chen, "The Influence of Voice on Pedagogical Agent's Persona and Recall Performance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64, 490–494* (SAGE: Publications, 2020).

[56] T. W. Morris and H. T. M. Chen, "Voice's Effect on Student Effort Ratings and Recall Performance While Under Cognitive

Stress," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65, 576–580* (SAGE: Publications, 2021).

[57] J. Leppink, F. Paas, C. P. van der Vleuten, T. van Gog, and J. J. van Merriënboer, "Development of an Instrument for Measuring Different Types of Cognitive Load," *Behavior Research Methods* 45 (2013): 1058–1072.

[58] J. D. Storey, "A Direct Approach to False Discovery Rates," *Journal of the Royal Statistical Society: Series B* 64 (2002): 479–498.

[59] J. M. L. Andres, M. M. T. Rodrigo, J. O. Sugay, et al., "An Exploratory Analysis of Confusion Among Students Using Newton's Playground," in *Proceedings of the 22nd International Conference on Computers in Education*, ed. C.-C. Liu (Asia-Pacific Society for Computers in Education, 2014), 55–64.

[60] Y. Jiang, L. Paquette, R. S. Baker, and J. Clarke-Midura, "Comparing Novice and Experienced Students Within Virtual Performance Assessments," in *Proceedings of the 9th International Educational Data Mining Society Conference* (International Educational Data Mining Society, 2015), 185–190.

[61] J. Ocumpaugh, M. O. San Pedro, H. Y. Lai, R. S. Baker, and F. Borgen, "Middle School Engagement With Mathematics Software and Later Interest and Self-Efficacy for STEM Careers," *Journal of Science Education and Technology* 25 (2016): 877–887.

[62] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)* 57, no. 1 (1995): 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[63] T. V. Perneger, "What's Wrong With Bonferroni Adjustments," *BMJ* 316, no. 7139 (1998): 1236–1238, https://doi.org/10.1136/bmj.316.7139.1236.