



Enhancing the Educational Potential of Online Movement Videos

System Development and Empirical Studies with TikTok Dance Challenges

Jules Brooks Blanchet

Dartmouth College
Hanover, New Hampshire, USA
jules.b.blanchet.gr@dartmouth.edu

Megan E. Hillis

Dartmouth College
Hanover, New Hampshire, USA
megan.e.hillis.gr@dartmouth.edu

Yeongji Lee

Psychological and Brain Sciences
Dartmouth College
Hanover, New Hampshire, USA
yeongji.lee.gr@dartmouth.edu

Qijia Shao

Hong Kong University of Science and
Technology
Hong Kong, China
qijia.shao@ust.hk

Xia Zhou

Computer Science
Columbia University
New York, New York, USA
xia@cs.columbia.edu

Devin Balkcom

Computer Science
Dartmouth College
Hanover, New Hampshire, USA
devin.balkcom@dartmouth.edu

David J. M. Kraemer

Department of Education
Dartmouth College
Hanover, New Hampshire, USA
david.j.m.kraemer@dartmouth.edu

Abstract

We hypothesize that online movement videos have untapped potential for teaching physical skills, and we developed a platform that automatically generates practice plans from raw TikTok dance videos. The practice plans teach one segment at a time using fading guidance and part-learning principles and are presented using a web-based interface featuring concurrent visual aids. Two user studies ($n=54$, $n=38$) were conducted. The first showed significant improvements in learning outcomes compared to standard tutorials, underscoring the importance of well-structured practice plans and offering nuanced insights into the design and effectiveness of visual aids. The second study found that segmentation and emoji-based dual-coding only benefit learning when integrated into a well-designed lesson structure. We provide a set of practical recommendations for enhancing online movement learning, focusing on the need for substantive part-learning activities and careful use of visual aids to prevent cognitive overload.

CCS Concepts

• **Human-centered computing** → **Empirical studies in interaction design**; **Empirical studies in HCI**; • **Applied computing** → *Performing arts*; *E-learning*.

Keywords

Human-Computer Interaction, User Experience Design, Embodied Interaction, Multimedia Learning, Motor Learning, Observational

Learning, Culturally Responsive Teaching, Part Learning, Fading Guidance, Dance, Dance Education, TikTok,

ACM Reference Format:

Jules Brooks Blanchet, Megan E. Hillis, Yeongji Lee, Qijia Shao, Xia Zhou, Devin Balkcom, and David J. M. Kraemer. 2025. Enhancing the Educational Potential of Online Movement Videos: System Development and Empirical Studies with TikTok Dance Challenges. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3706598.3714062>

1 Introduction

Expressive movement, from dance to fitness routines, is a core aspect of the human experience. Movement is showcased in a wide range of online videos, but despite the abundance of inspiring content the educational utility of these videos often remains under-explored in their standard format. TikTok dance challenges, a cultural phenomenon that blends skill acquisition with social engagement, serve as an ideal model for examining effective online learning environments within the domain of physical skill learning.

Our research is twofold. First, we develop a system designed to enhance the educational delivery of movement videos and conduct a study to evaluate the learning outcomes and gather user feedback. Second, by analyzing how the features of user-created dance tutorial videos contribute to learning outcomes, we use this platform as a case study to understand broader principles impacting the creation and presentation of educational movement content.

In study 1, we develop a web platform for dance learning driven by lesson plans that are automatically constructed from in-the-wild dance videos. Our vision is to expand access to quality motion instruction for those who may not be able to afford a human coach. Existing dance teaching systems typically require complex equipment and manual content creation, thus limiting scalability and adaptability to social trends (see Section 2.2). To enable maximum access, our offering minimizes equipment requirements, supporting



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '25, Yokohama, Japan*

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3714062>

devices that are common in people homes (e.g. laptops or tablets with webcams), and generates learning experiences automatically from raw motion videos.

Our application embeds a software pipeline that takes dance videos as input, extracts human pose representations, derives meaningful features such as key frames and 8-count segmentations, and finally compiles a practice plan. These auto-generated practice plans are presented in a scaffolded web interface to enable accessible guided learning, implementing features such as *incremental part learning*, which breaks down complex movement sequences into smaller, manageable segments for step-by-step mastery, and *fading guidance*, which gradually reduces instructional support as learners gain proficiency [11, 37]. The web app guides users through the lesson plans step-by-step and employs visual aids such as motion overlays.

In our first user study, we evaluate the visual aids and the system as a whole, collecting quantitative and qualitative data to validate the theoretical predictions underpinning our design, understand the outcomes resulting from our specific feature implementations, and formulate a generalizable set of recommendations for future creators of motion training systems.

Study 2 was inspired by qualitative user feedback collected under study 1, which suggested that emoji-segment labels present in the creator-authored videos of the control condition could be beneficial to learning—a proposition subject to mixed theoretical predictions. On the TikTok platform some creators offer dance challenge tutorials, which are typically formatted with rectangles overlaid on the video containing emoji and text labels of the dance moves. These label overlays are time-synced to the video, changing background color when the corresponding dance move is being performed (Fig. 1). The dance tutorial videos are popular – for example, as of July 2023, videos with the *#dancetutorial* hashtag have 18.7B views as compared to 53.8B for *#dancechallenge*. This emergent style of dance learning suggests user interface features that hold promise for use in automatically generated learning experiences, but which have not been empirically tested. Using our system as a technical probe, we analyze how the segmentation and emoji-segment label features influence learning outcomes and generate insights on how specific elements could be applied in automatically generated learning scenarios.

Contributions

This work makes the following contributions to the field of online movement learning systems:

- **Artifact Contribution:** We present a novel system that automatically generates practice lesson plans based on motor learning theory using online, “in the wild” video content.
- **Theoretical Contribution:** We validate core motor learning techniques such as segmentation, incremental part-learning, and fading guidance in the context of online video-based dance learning. We provide insights into how emoji segment labels function as dual-coding tools in dance instruction.
- **Empirical Contribution:** Based on quantitative and qualitative results from two user studies, we offer actionable

design insights, including recommendations on effective segmentation, the use of visual aids, and strategies for managing cognitive load.

2 Related Work

This section provides an overview of prior work in two main areas: (1) theoretical foundations from motor learning theory that inform the design of our system and (2) dance teaching systems developed within the Human-Computer Interaction (HCI) community. We compare our approach to prior technological systems for dance instruction and outline key insights from motor learning literature that guide the development of our automatically generated practice plans.

2.1 Motor Learning Theory

Motor learning theory describes how learners progress through different stages of motor skill learning, from cognitive to associative to autonomous ([10, 20] – although other models have been proposed; see [32]). Our system focuses on the consolidation process of moving from the cognitive stage to the associative stage.

Motor learning theory highlights the importance of feedback during skill acquisition, which can be explicit (system-provided performance evaluations) or implicit (helping learners recognize their own errors). Augmented visual and multimodal feedback is especially effective in early learning stages but may impede knowledge transfer later on due to the guidance hypothesis, which suggests that excessive reliance on augmented feedback can prevent learners from developing intrinsic error-detection and correction mechanisms, ultimately reducing their ability to perform the skill independently [36, 37]. In TikTok dance challenges, music provides an essential source of implicit & concurrent auditory feedback. In the visual modality, overlaying the target and realized motion via a skeletal illustration has been shown to be particularly effective [1, 21, 40].

Motor learning literature supports the use of part learning for complex skills like dance routines [11]. Incremental part learning, where components are mastered one by one, aligns with professional dancers’ practice routines and forms the basis of our system’s approach to lesson design [28]

Wulf and Lewthwaite [45] propose the *OPTIMAL* framework for optimizing motor learning, which emphasizes autonomy support, encouraging positive performance expectations, and adopting an external focus of attention. Self-controlled practice has been shown to result in better learning than externally-imposed practice [46]. In general, learners tend to perform and learn better when they are offered choice and encouraged to exhibit agency in their learning process [45]. Our system incorporates motor learning optimization principles by supporting user autonomy and offering choices in practice routines, both of which have been shown to enhance skill acquisition [45].

2.2 Dance Teaching Systems

Zhou et al. [48] surveys some of the substantial HCI work in dance, examining empirical studies, choreographic tools, motion analysis

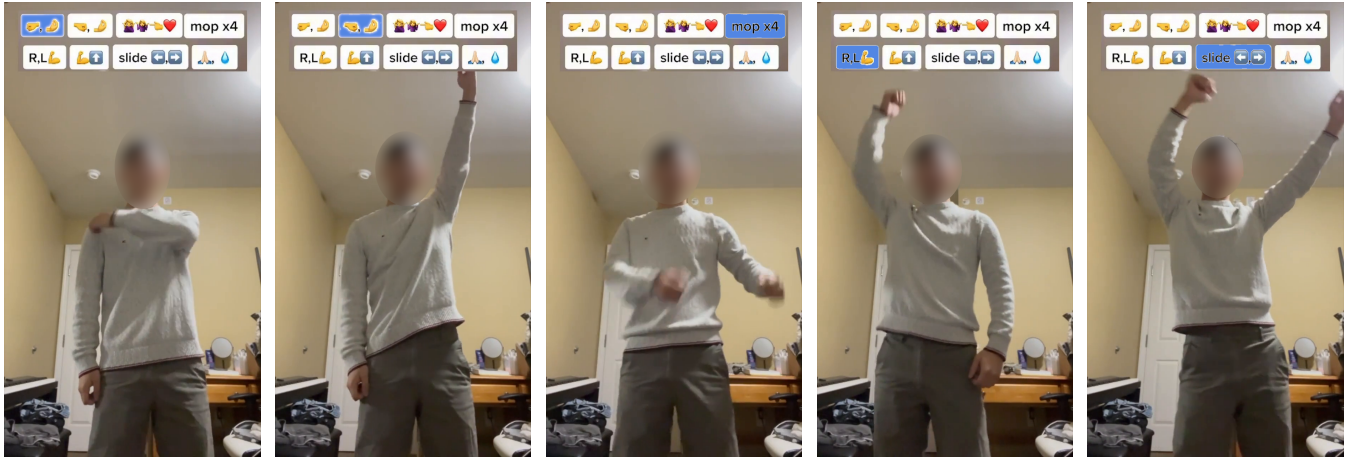


Figure 1: TikTok dance tutorial video format. The videos are overlaid with boxes representing a segmentation of the dance, which contain emoji labels of the dance moves. The boxes change background color in as the dance progresses.

techniques, dance performance augmentation, and dataset contributions. Raheb et al. [25] survey dance learning systems specifically. We summarize a selection of proposed systems in Table 1.

Several dance teaching systems have been proposed, with various designs related to instructional guidance, required equipment, content source, and feedback. With YouMove, Anderson et al. [1] propose a generalized motion teaching system using an augmented reality mirror. Their system is both an authoring tool for dance experts to record & annotate instructional motions and a teaching tool that guides learners through five stages of learning - *demonstration*, *posture guide*, *movement guide*, *mirror guide*, and *on your own*. While YouMove uses expert-authored content and proprietary hardware, our system automates content generation from existing videos and is designed to scale using widely available devices.

The salsa VR and and Greek dance applications by [15, 35] offer specialized dance learning experiences but are limited by their reliance on specific hardware setups and a lack of broader applicability to other dance styles. In contrast, our system is domain-general and designed for scalable, minimalist hardware.

With SyncUp, Zhou et al. [50] describe a tool to support the practice of synchronized dance groups, assisting them by automatically detecting and highlighting periods of unsynchronized movement on video recordings, enabling faster iterative practice sessions. This approach does not offer support for beginners in the process of initially learning the dance. E-Ballet [41] experiments with feedback for a set of ballet movements in an e-learning setting, using a Kinect sensor. Trajkova and Cafaro [42] take a similar approach for comparing visual vs auditory and corrective vs value forms of feedback. The authors make use of emojis to represent visual value (e.g. performance score) feedback, though this is substantially different from our user study 2 investigation of emojis as symbolic representations of dance moves.

There are some commercial digital products that provide a dance learning experience, generally falling into one of two categories: games designed for entertainment, such as Just Dance, and products designed for dance learning, such as STEEZY Studio and learn-to-dance.com. These products operate on manually-created content,

adding to their expense and restricting their ability to scale to the expansive array of dance styles and skills that exist.

Imitation of dance movements seen in a video is a basic case of using technology to learn motor skills by example. This experience can be augmented with additional visual or multimodal effects to support the learning experience, as discussed in *Motor Learning Theory*. In their tool for synchronized dance practice support, Zhou et al. [50] describe two visual interfaces for communicating the synchrony of dance moves: a heatmap overlay that highlights discrepancy in poses, and a color-coded timeline that communicates periods of temporal misalignment. Beyond the AR mirror hardware, Anderson et al. [1] has two more visual aids: posture guide, which presents a skeleton overlay for static poses, and movement guide, which presents a moving skeleton overlay alongside “ribbon” that cue upcoming movements. Hu et al. [13] summarize four existing visualizations for human motion data (2D motion map, action synopsis, motion belts, and representative video clips) and propose a 5th, “Motion Track”, which embeds keyframes into a 2D space and draws a curve along that space (“motion track”) to represent the motion. This method is effective for summarizing and differentiating motion sequences, but may not be suitable for learning as the representation lacks temporal information.

Clarke et al. [8] describe a system for video-based motion learning whereby playback of a demonstration video is slowed, paused, sped-up, or rewound in order to synchronize the video with the user’s current pose, with a skeleton overlay and hand travel-direction cues. Tsuchida et al. [43] use deepfake technology to synthesize videos of the learner performing the source dance but found that it had a neutral to slightly-negative impact on learning outcomes. This exemplifies the importance of considering and validating pedagogical intent when applying new technologies to digital learning platforms. Finally, Singh et al. [38] propose a web-based system that allows users to add time-attached text, ink, and video notes to recorded videos.

While many existing dance teaching systems focus on augmented feedback mechanisms such as motion overlays, heatmaps,

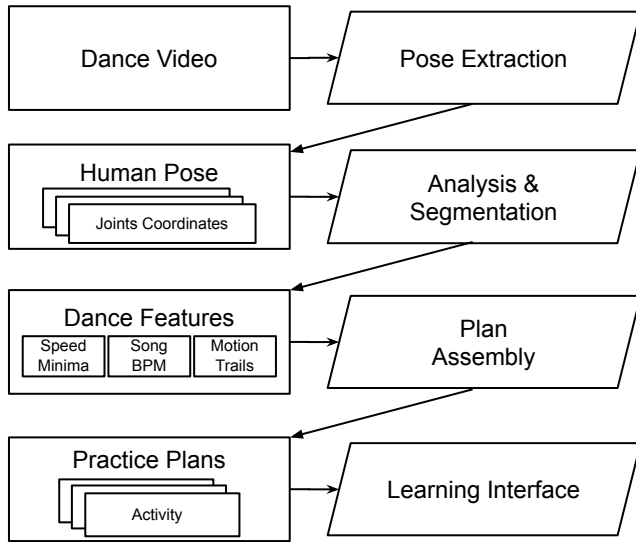


Figure 2: The Practice Plan Compilation Pipeline

and video playback controls, few integrate structured motor learning techniques such as incremental part learning and fading guidance. Our system differs in that it explicitly structures learning experiences using segmentation, guided progression, and adaptive scaffolding, aligning with well-established principles of motor skill acquisition rather than relying solely on real-time feedback or self-guided video-based learning.

3 System Engineering

Our system is composed of two components: a practice plan compiler, which accepts a dance video from the Internet and generates a lesson plan, and a user interface, which presents the instructions, content, and visual aids embedded in the lesson to the learner, guiding them through the dance.

Practice Plan Compiler. The practice plan compilation pipeline consists of multiple stages of data processing where the last stage produces practice plans that are later consumed by the learning interface (see Fig. 2). First, we use the MediaPipe framework [19] to extract a skeletal pose of the dancer in each frame. The resulting time sequence data contains normalized x and y positions of 33 skeletal landmarks within the video frame, as well as estimated 3D coordinates of those landmarks in the world frame.

In the motion analysis stage, multiple aspects of the dance are analyzed to produce a higher-level representation of the motion. First, landmark data is re-referenced to the torso center and normalized by the torso length. This allows subsequent analysis to be scale, distance, and frame-position independent. We compute: (a) the speed minima of a subset of joints in 3D space, (b) tempo of the accompanying music, and (c) simplified motion trails of the hands. The decomposition algorithm then chooses a decomposition strategy for the dance based on these features, preferring to use a tempo-based segmentation of 4-beat bars but falling back to a speed-minima segmentation in the case of an undetectable tempo –

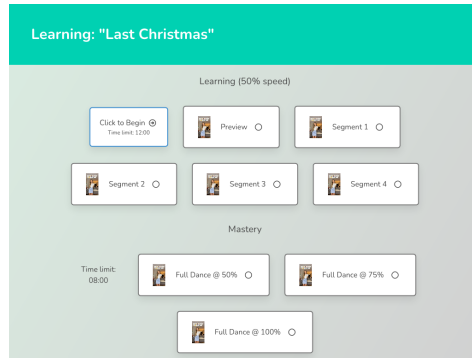
though the tempo was detectable for all videos used for this paper’s user studies.

Finally, the practice-plan assembly stage generates a sequence of learning activities—i.e. a ‘practice plan’—using a structured methodology based on predefined instructional design principles. Each of these activities consists of one or more learning steps, with the structure of the practice plan tailored to the specific research objectives. The details of this methodology, including the specific sequencing of instructional activities, are described in Study 1 and Study 2, where we outline how the practice plans were designed for each study. In study 1, we tailored the practice plan to optimize learning outcomes, thereby establishing a benchmark for evaluating our system against the instructional capabilities of TikTok tutorial videos. For study 2, the practice plan was stripped down, removing potentially confounding features of our system in order to compare the effects of segmentation and emoji-labeling on learning outcomes in a setting comparable to virtual learning from video.

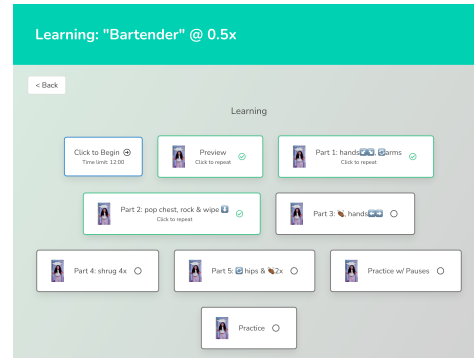
The most computationally intensive step in the compilation process is skeletal pose inference from video, which takes approximately 38.5 seconds for a 15-second video on an M2 MacBook Air. Tempo analysis of the audio track is a separate step, adding around 2 seconds per video. The remaining processing steps, including segmentation and practice plan assembly, are lightweight and complete in under 1 second, ensuring that most of the computational overhead is concentrated in the initial video analysis stages.

User Interface. The second component of our system is the learning interface, a web application that guides the user through the automatically generated lessons. Designed with insights from *Motor Learning Theory*, the interface follows Nielsen’s 10 UI design principles [24], which are widely recognized in the HCI community as foundational heuristics for evaluating and guiding the development of intuitive, efficient, and user-friendly interfaces, emphasizing usability, error prevention, and clear system communication. Upon visiting the site, users are greeted with a lesson menu screen (Fig. 3a) that visualizes the learning path ahead of them, allows the user to navigate the learning process, and shows the user’s progress in accordance with Nielsen’s *visibility of system status* principle. Although learning activities are presented in the intended order, users are free to select activities in any order, permitting user control and freedom as suggested by Sigrist et al. [37] and reflected in Nielsen’s *user control and freedom* principle. According to the experimental needs, these activities can be labeled with generic titles (such as Fig. 3a for study 1) or with richer semantic titles (such as the segmented-emoji conditions in study 2, Fig. 3b).

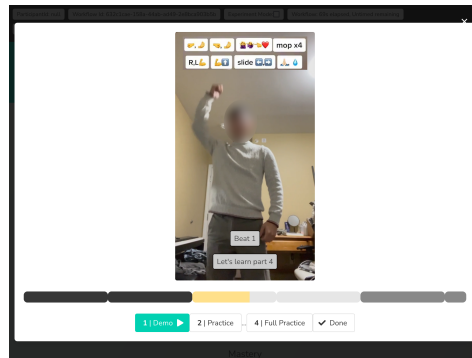
Our practice plan provides structure, and the user is free to practice different segments of the dance with whatever degree of feedback they desire. Our instructions empower users to decide when they feel confident enough to progress, encouraging them to repeat an activity until they personally feel proficient. The availability of these choices offers users agency over their learning process, fitting within the OPTIMAL framework [45] – a theory emphasizing autonomy, enhanced expectancies, and external focus of attention as key factors in optimizing motor learning, which aligns with our goal of designing instructional experiences that foster engagement and effective skill acquisition.



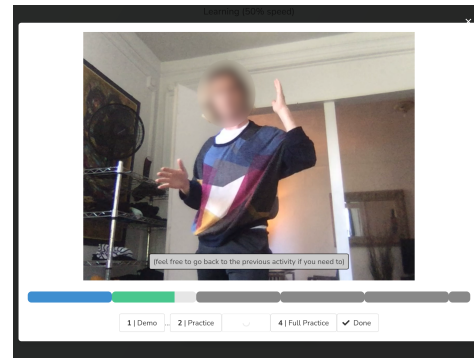
(a) Practice Plan Menu (Study 1, System with Skeleton Overlay Condition)



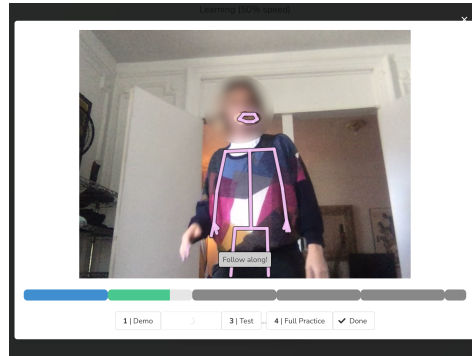
(b) Practice Plan Menu (Study 2, Emoji-Segment Condition)



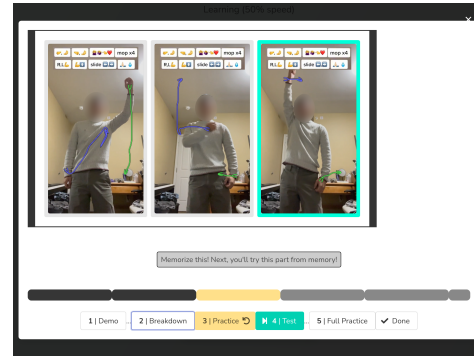
(c) Learning Activity: Demo Video



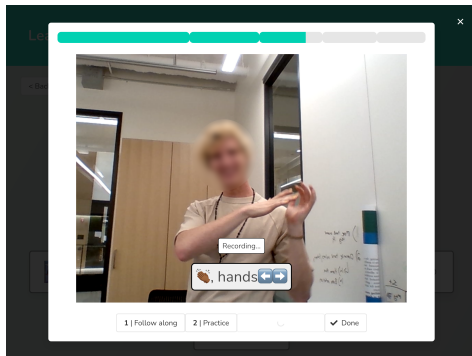
(d) Learning Activity: Virtual Mirror



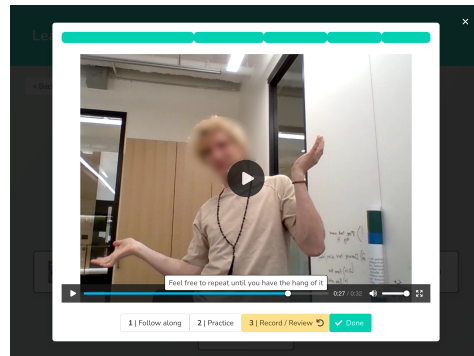
(e) Learning Activity: Skeleton Overlay



(f) Learning Activity: Sheet Motion



(g) Learning Activity: Record



(h) Learning Activity: Review

Figure 3: User Interface. The functionalities shown here can be incorporated into practice plans as needed according to learning or experimental needs. Behavior of these screens is described in the *User Interface*.

Table 1: Attributes of dance teaching systems in the literature.

System	Instructional Guidance	Equipment Setup	Content	Feedback	Evaluation
<i>YouMove</i> [1]	Guided Activity Sequence	Kinect sensor & AR Mirror	User-authored, motion-general	Visual, explicit & implicit	n=8 comparing system with traditional video instruction
<i>Syncup</i> [50]	No specific guidance.	RGB Webcam & Monitor	User-authored, Synchronized troupe dances	Visual, explicit & implicit	n=9 user study, evaluating dance groups' impressions of the system.
<i>Salsa dance learning</i> [35]	Guided	HTC Vive with extra trackers	Manually-created, 2 specific salsa partner-dances	Visual, Explicit (score)	n=40 comparing dancers vs non-dancers use of the system
<i>A Game-like Application for Dance Learning</i> [15]	Guided	Kinect Sensor(s) & Monitor	Manually Created	Visual, explicit	n=18 user study, examining participants' score changes after using the system.
<i>E-Ballet</i> [41]	Guided	Kinect sensor & Monitor	Manually created	Visual & verbal, explicit & corrective feedback, using 'Wizard of oz' approach	n=16 user study, gathering user impressions of feedback through interviews.
<i>Super Mirror</i> [21]	No specific guidance	Kinect sensor & Monitor	Manually Created, Static Ballet Poses	Visual, Implicit	n=5, interview-style, gauging user impressions of the system
<i>WhoLoDance Tools</i> [26, 31]	Not Guided	Optical motion capture, Monitor or Microsoft Hololens	Manually Created, 4 styles of dance	Visual, Audible, Text; Implicit, Explicit	
<i>Delay Mirror</i> [23]	Not Guided	RGB Webcam & Projector	User-authored, Delayed video stream	No augmented feedback	n=8 participants, evaluating utility in the context of a dance class
<i>MoveOn</i> [29]	Guided	RGB Webcam & Monitor	User-authored, motion-general	No augmented feedback	Series of 3 workshops (n=4, n=6, n=6) examining decomposition strategies
<i>HereAndNow</i> [49]	Not Guided	Kinect Sensor & AR Mirror, Logitech Presentation Remote	User-authored, motion-general	Visual, Implicit	Series of 3 workshops + Survey (n=13); Interviews with expert dancer, choreographer, and digital media artist (n=3)
<i>LearnThatDance</i> (Present work) [2–4]	Guided Activity Sequence with User Choice	RGB Webcam & Monitor	Automatically generated, motion-general	Visual, implicit	Two studies: n=54 comparing system to TikTok video tutorials, n=38 examining TikTok tutorial video format

Figures 3c to 3f show the learning activity interface, which appears after a user clicks on one of the activities on the lesson menu screen. Users can interact with the lesson's steps—play, repeat, and navigate—using buttons located at the bottom of the interface. A progress bar above the controls visually segments and color-codes dance movements, aligning with choreography segments for straightforward tracking. The central area displays content such as a demonstration video (Fig. 3c); a webcam feed with a *skeleton overlay* (Fig. 3e); a *sheet motion* presentation of key frames (Fig. 3f); or a *review/record* display, offering the users the ability to record and review their performances, thereby supporting a learning technique common among professional dancers [28] – see Fig. 3g and Fig. 3h.

Two of the above forms of content are visual aids that go beyond what is available in typical learning-from-video scenarios: *skeleton overlay* and *sheet motion*. *Skeleton overlay* mode (Fig. 3e) displays a 2D stick figure over the user's webcam feed. The stick figure reflects the demonstrator's pose, guiding users to mimic these positions with their bodies. This visual aid provides implicit, concurrent feedback [37] that delivers a portion of the choreographic information available in the reference video, thus scaffolding the progression from mimicking the reference video to performing the dance from memory. Variants of this approach has been implemented in prior work [1, 14, 21, 34] and have been shown to be effective in supporting movement learning.

Sheet motion mode (Fig. 3f) arranges still images from each beat's onset, organizing them in rows that match the music's rhythm. Images align with musical bars, with four per row to match the $\frac{4}{4}$ time signature of the dances used in this study. As the dance progresses, the still images corresponding to the current time are highlighted. To cue learners as to the motion that occurs between the still images, arrows indicating the path of motion of the hands are drawn atop the still images. This presentation mode is a combination of the *motion belt* and *motion cues* movement summarizing techniques described in Li et al. [18], and, similar to *skeleton overlay*, is intended as a scaffold between presenting the full information of the reference video and having learners perform the dance entirely from memory.

The system supports the ability for text cues to appear at specific times in the video, which was used in the emoji conditions of study 2, as visible in Fig. 3g. The interface also implements a *performance upload* functionality, enabling users to record and upload a video of themselves performing the dance while accompanied by the music.

4 Methods

We conducted two studies of our system using a similar methodology but with different foci of inquiry. Our research process started with a focus on operationalizing and testing insights from motor learning literature in the context of an automated system for motion skill teaching.

Study 1 (Evaluation of Interactive Dance Teaching System). Evaluates the performance of two variants of our dance teaching system compared to self-guided learning from TikTok tutorial videos, with the goals of determining whether our system enhances learning-from-video and uncovering how different features of our system affect the learning process. The study used a 3-condition design per Table 4.

Intrigued by the qualitative results of this first study, we then conducted a second study to investigate the educational potential of design features found in TikTok dance video tutorials.

Study 2 (Investigation of Features of TikTok Dance Tutorials). Leverages our system as an experimental tool to examine the emoji segment labels and dance segmentation features that commonly present in TikTok dance-challenge tutorial videos and determine their effect on learning outcomes.

Evaluation of learning outcomes and qualitative feedback are discussed later in this section.

Approval. The study procedure was approved by the Committee for the Protection of Human Subject at Dartmouth College. Informed consent was obtained from all participants prior to their participation.

Participants. Both studies were conducted online with participants recruited from our local community. 54 participants were recruited for study 1, (36 female, mean age = 18.67 years) and 38 were recruited for study 2 (20 female, mean age 22.26 years). Across both studies, the majority of participants (66.7% in study 1, 68.4% in study 2) reported no prior dance training, and the experience of those who did ranged from “a few lessons” to 15 years of dance

experience. To control for participants' varied levels of dance experience, we employed a within-subject procedure (described below) and included a participant random intercept in our linear mixed model analysis (as described in each of the study result sections). After each experiment, no participants reported that they had ever previously encountered the specific TikTok challenges used in the study. No participants from study 1 also participated in study 2.

Procedure. Participants learned different dances in three (study 1) or four (study 2) lesson conditions, presented in a randomized order with counterbalanced dance-condition assignments. Each dance video features a single dancer and lasts between 13.97 and 18.15 seconds—see Table 2 to for the list of dances that were used. The dance videos used in this study were selected by manually reviewing videos under the *#dancetutorial* tag on TikTok. Selection criteria included (1) featuring a single dancer, (2) ensuring the entire body was visible in the frame to allow for accurate pose extraction, and (3) having an approximate 15-second duration.

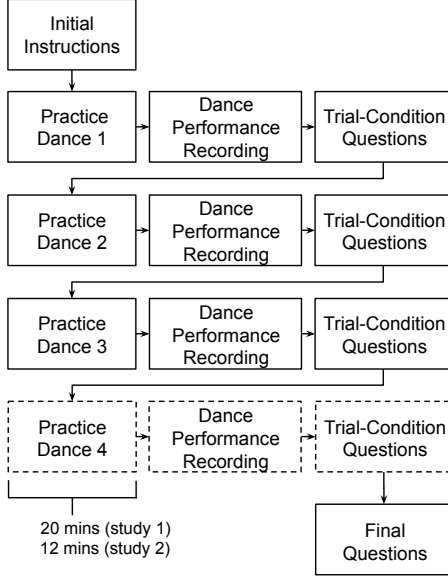
In each lesson, participants interacted with our system under a practice plan to according to the assigned experimental condition, as described in Section 5: *Practice Plan* and Section 6: *Practice Plan*. Participants were free to navigate the app and engage in practice as they saw fit within a fixed amount of practice time (20 minutes total for study 1 and 12 minutes for study 2). After each lesson, participants used the platform to record a video of themselves performing the dance they just learned. Participants then rated dance difficulty and system helpfulness, and answered open-ended questions regarding what was helpful and what could be changed or improved in the lesson. Additionally, after the final lesson, participants completed a longer questionnaire which gave them a chance to comment about the system as a whole and included the system usability scale (SUS) assessment [5]. We analyze both performance videos and the post-lesson questionnaire responses to compare the conditions in each study. The SUS results are not considered here, as the participants completed this assessment only once at the end of the study and it was unclear which conditions their assessment results would be applicable to.

Analysis: Performance Accuracy Score. Similarity between the participant's video and the reference TikTok video was chosen as the main criterion for performance accuracy. Based on the pose similarity analysis algorithm proposed by [50], we utilize the skeleton information extracted from MediaPipe [19] to automatically score the accuracy of each learners' uploaded performances. Our target dances emphasize upper body movements; therefore, we chose 8 key points from the upper body and compute 8 key vectors, as illustrated in Fig. 5, which capture the upper-body movement of the dancers. Although Mediapipe computes landmarks in the lower body, dancers' lower bodies are frequently out of the video frame in TikTok dances; therefore, we omitted lower-body vectors from our accuracy scoring.

Comparison of joint vectors between the learner and expert is affected by the variance in body proportions and camera recording distances. To address this issue, we normalize the key vectors into unit vectors which provide directional information for each body part that is invariant to the above factors. Our method assumes that the user and reference dancers are performing at a similar orientation relative to the camera. Given the nature of TikTok dance

Table 2: The dances used in our study. Likes are as of 2023-07-04.

Id	Date	Creator	Choreographer	VideoId	Likes	Learned in
A	2020-12-20	@helenpenggg	@ruby.bauer	6908526401658391813	179.9k	Studies 1 and 2
B	2020-10-18	@phoebe.mulyana	@leilanigreen	6884913446505254145	402.5k	Studies 1 and 2
C	2021-01-24	@helenpenggg	@joitie04	6921519498767928581	23.1k	Studies 1 and 2
D	2021-05-04	@koristutorials061	@sauxyyjay	6958501156406578438	223.5k	Study 2 only

**Figure 4: User Study Flow. The participants attempted three (study 1) or four (study 2) trials, one dance for each condition.**

challenges, this assumption is met in practice, as dancers mostly face their recording device throughout TikTok dance, including the ones we used for our user studies.

In each frame, we compute the absolute difference between the corresponding unit vectors of the learner and the expert (range: $[0, 2]$, with 0 indicating perfect alignment), and then sum them up as the per-frame error. The overall error is calculated as the average of all frames of the dance. Finally, we rescale the score into the range of $[0, 5]$, where 0 denotes the poorest performance and 5 represents the best performance. This normalized score serves as the final performance accuracy.

We tested our automatic scoring system by comparing its output to human rated scores. For each study, three human raters were instructed to rate how similar the given participant’s dance was to the teacher’s dance. The human raters were recruited from our local university. Among the six human raters, two had no dance training, while the other four had varying levels of experience ranging from a few months in 2nd grade to seven years in elementary school. For each study, segment pairs (participant video segment, reference video segment) were shown to three raters who then scored the dance segments on a scale of 1 (least similar) to 3 (very similar). The scores for each performance were calculated by taking the arithmetic mean among all segments and then again among all

raters. In both study 1 and 2, the three raters demonstrated high inter-rater reliability with a Krippendorff’s alpha of $\alpha = .745$ and $\alpha = 0.601$ respectively.

In order to compare the human and automatic scores, the human scores were rescaled to align with the $[0, 5]$ range of the automatic score system. There was a high correlation between human and automatic scores (Pearson’s $R = .93$, $n = 128$, $p < .001$) as well as in their rank orders (Spearman’s $r_s = .90$, $n = 128$, $p < .001$ – Fig. 5a) in study 1. This was the case in study 2 as well (Pearson’s $r = .93$, $n = 147$, $p < .001$, Spearman’s $r_s = .90$, $n = 147$, $p < .001$ – Fig. 5b). In addition to this robust correlation, the human ratings displayed a ceiling effect, suggesting that automatic scores measure more variance than human ratings. Despite this observation, the substantial correlation implies that our automatic scoring approach mimics human similarity ratings, indicating its usefulness in gauging the similarity of dance performances.

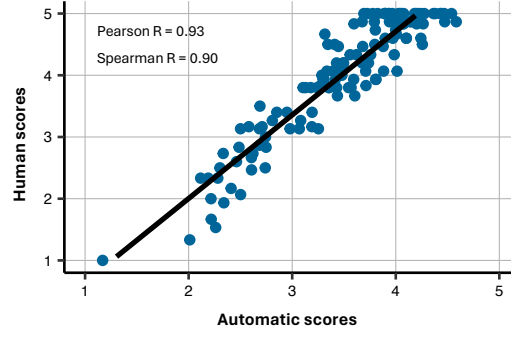
Analysis: Dance Complexity & Difficulty. Dances vary in speed, complexity, and performance difficulty. These factors could modulate the effectiveness of different learning features. To measure this, we asked participants to rate the level of difficulty they experienced learning each dance after completing the corresponding lesson. Three dances were used in study 1 (dances A, B, and C) and all four dances were used in study 2.

While self-reports can capture individual variation in perceived difficulty, we also developed the *motion complexity* metric to provide a more objective basis for comparison and analysis across the selected dances. The metric is calculated by taking the mean velocity, acceleration, and distance across all skeletal joints, and then z-scoring these motion parameters as well as the tempo (in beats-per-minute) relative to the full dataset of dances. Finally, we combine the motion parameter z-scores to form a scalar complexity value for each dance. From the composite of these metrics, dances B and C were assessed as the most complex dances, and dance A as the least complicated, as shown in Table 3.

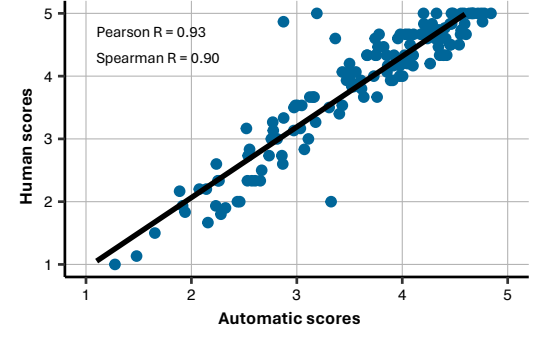
Defining a metric that attempts to capture the difficulty of performing a motion is challenging, due to factors such as the highly individualized nature of physical abilities and learning rates, the multidimensional aspects of motion including coordination, balance, and flexibility, and the subjective perception of difficulty which varies greatly among individuals. For this reason, we designed our metric not as a definitive measure of difficulty but as a comparative tool to objectively assess movement complexity across dances, adding nuance to our analysis of learning aids’ effectiveness. Our approach is relatively simple, taking into account the amount of movement but not the significance or relative challenge of movements; more sophisticated approaches exist [39, 47].



Figure 5: Automatic performance rating. Our method focuses on upper body vectors.



(a) Study 1



(b) Study 2

Figure 6: Correlation of Automatic and Human scores

Table 3: Dance Difficulty Metrics

Id	Duration	Complexity (Z-Score)	Complexity per Second	Mean Rating (Study 1) (out of 5)	Mean Rating (Study 2) (out of 10)
A	15.1s	-1.163	-0.07	2.49	4.34
B	18.2s	0.551	0.03	3.29	5.68
C	14.0s	0.832	0.06	3.56	6.16
D	14.9s	-0.220	-0.01	N/A	6.00

Analysis: Qualitative Feedback Synthesis. After each trial in both studies, participants were asked two open-ended questions: “What did you find helpful about this learning experience?” and “What would you change/improve about this learning experience?” This allowed us to gather insights into user perceptions of different system conditions and elicit suggestions for future improvements in a semi-structured manner. Three of the authors performed a thematic analysis of the user responses in two stages, using the taguette tool [27]. First, each author independently coded the responses, developing codebooks that grouped key sentiments or ideas reflected in the participants’ feedback. After this initial coding, the coders met to collaboratively reconcile their codebooks, resolving discrepancies and producing a final, unified codebook. Finally, one author recoded all the responses using the reconciled codebook, and the frequency of each code within each lesson type was tabulated to identify patterns across conditions.

5 Study 1: Evaluation of Dance Teaching System

In this section, we present the findings from Study 1, in which we evaluate the effectiveness of our interactive dance teaching system in comparison to TikTok tutorials. The goal was to determine whether our system enhances learning outcomes and to investigate how different system features influence the learning process.

5.1 Practice Plan

The practice plan for study 1 was designed to optimize learning outcomes. The plan starts with a learning phase, to introduce a dance

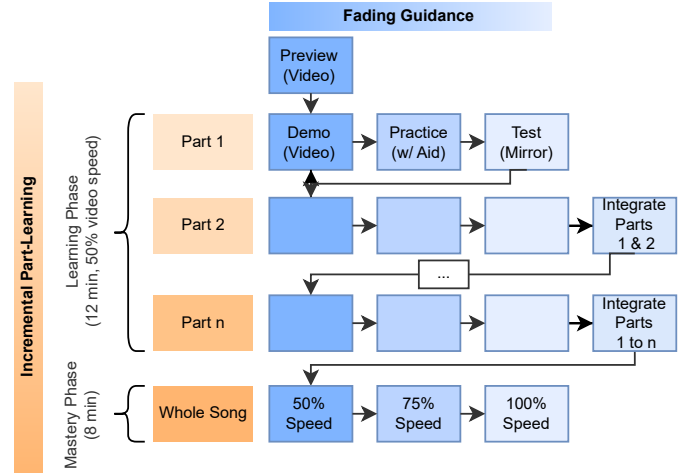


Figure 7: Practice Plan for Study 1 (non-control conditions). The structure of the practice plan incorporates two techniques drawn from motor learning literature: fading guidance and incremental part-learning.

to a user who has never seen it before. The first activity is a preview of the dance, in accordance with Mayer [22]’s *pre-training* principle, which states that activities performed prior to a challenging main task can assist the user in managing essential processing, lowering

Table 4: Study 1 Conditions: Two system variants compared with a control condition.

Condition	Lesson Format	Visual Aid
Control: Tutorial Video	Single Activity with Video Controls	Creator-embedded Emojis
System: Sheet Motion	Autogenerated Lesson	Sheet Motion Display
System: Skeleton Overlay	Autogenerated Lesson	Overlaid Skeleton

the cognitive load of the main task to a more suitable level. All if the remaining activities in the learning phase occur at half speed. The preview is followed by a sequence of activities that incrementally teach the dance segment-by-segment in accordance with Mayer’s *segmentation* principle, following a part-learning approach [11] as shown in Fig. 7. As described in *Practice Plan Compiler*, this segmentation is generated automatically by the practice plan compiler. For each of these activities, the user performs a set of steps with progressively reduced guidance, as follows: (1) a video demonstration of the segment (Fig. 3c), (2) a practice step with either a skeletal overlay (Fig. 3e) or sheet motion (Fig. 3f) learning aids, (3) a test step, in which the user sees their webcam feed and performs the segment from memory (akin to a ‘Virtual Mirror’, Fig. 3d), (4) an integration step, in which the user performs the entire dance up until the end of the newly learned segment. Visual aids are provided in this step to support the user in recalling previously learned segments. This progressive increase in task difficulty and removal of feedback support is informed by studies on motor learning which have found that concurrent visual feedback is helpful while initially learning a task (the cognitive stage, [10]), but that supports should be removed as the user becomes familiar with the motion (the associative stage) [37].

By the end of the learning phase, the entire dance has been incrementally introduced to the user. The compiler then adds a mastery stage, consisting of a series of activities that prompt the learner to practice the entire dance at increasing speeds (0.5x, 0.75x, and 1x). Each of these mastery activities follows the same demo-practice-perform sequence as in the learning phase. Practice plans are stored in a JSON format for use in the learning interface.

The TikTok videos we chose had time-synchronized symbolic emojis embedded by the creator, which were left visible in the control condition (as in Fig. 3c) but were obscured in the experimental conditions. In the experimental conditions, users were provided auto-generated practice plans as previously described. In control condition we used a manually-created practice plan with a single ‘free practice’ activity which played the demonstration video, with native video controls enabled (such as pause, play, seek, change playback speed). See Table 4 for a breakdown of the experimental conditions.

Thus, each participant gave feedback on each of the three experimental conditions, learning a different dance each time.

5.2 Results

We fitted a linear mixed effect model (LMM) to test the effect of lesson type on participants’ performance accuracy, as evaluated by the automatic system. The model included lesson type and dance as fixed effects, along with their interaction, with participants as a random intercept. The three different lessons (traditional *t* tutorial,

auto-generated lesson with skeleton aid, auto-generated lesson with sheet motion aid) were dummy-coded as a three-level factor, and the three different dances (dance A, B, and C) were coded using deviation coding [44]. This structure allows us to evaluate the mean of each lesson directly compared to the mean across all dances. The Levene test and visual inspection of residual plots revealed that the assumption of normality and homogeneity of variance was not violated ($F_{8,121} = 1.231, p = 0.237$). A conditional explanatory power of $R^2 = .681$ indicated that approximately 68.1% of the variance in automatic score was explained by this model. Estimated marginal means of participants’ performance accuracy by lesson type, lesson type within each dance, and dance—see Fig. 8. The fixed effects analysis showed a significant main effect for lesson type ($X^2(2, N = 130) = 7.922, p = 0.019$) and a significant interaction between lesson type and dance ($X^2(4, N = 130) = 11.779, p = 0.019$). This indicates that participants’ performance differs by lesson type and that this effect may depend on which dance they are learning.

Post-hoc Tukey HSD tests identified significant performance differences between lesson types following the LMM analysis. Results showed participants performed better in the auto-generated lessons with skeleton aid compared to those in the control condition using *t* tutorial videos ($t = -2.597, p = 0.030$). However, no significant difference was observed between auto-generated lessons with sheet motion aid and the *t* tutorials ($t = -0.360, p = 0.931$; see Fig. 8 left). Despite the varying complexity of the dances, no significant main effect of dance was found, ($X^2(2, N = 130) = 1.474, p = 0.478$) suggesting participants performed equally across the three dances (see Fig. 8 right).

Additionally, we compared performance between the lesson types within each dance to assess how the dances influence the effect of lesson type. This analysis revealed that in dance A (the least complex of the dances, see Table 3), users performed better with auto-generated lesson with sheet motion aid as compared to Tiktok tutorial ($t = -2.546, p = 0.033$), see Fig. 8 center. Conversely, in dance C (the most complex dance), users performed better with auto-generated lesson with skeleton aid than Tiktok tutorial ($t = -3.050, p = 0.008$). Moreover, comparisons within auto-generated lessons revealed a superior performance with skeleton aid over sheet motion aid ($t = -2.704, p = 0.022$). For dance B, there was no significant difference in users’ performance across lesson types. These findings indicate that users either outperformed or matched their performance with auto-generated lessons relative to the TikTok tutorials across the majority of dances.

We also conducted an additional LMM analysis to assess how users’ perception on helpfulness and difficulty of a lesson is different between the lesson types.

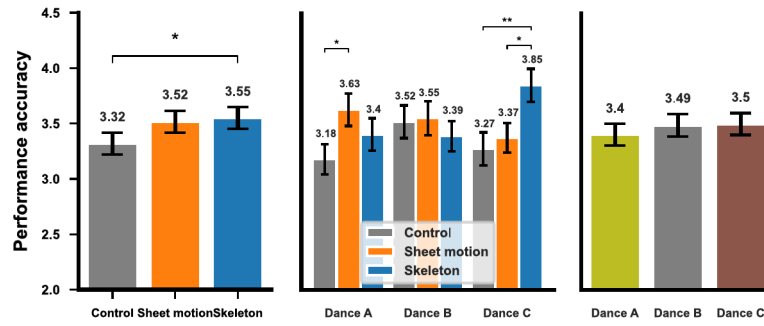


Figure 8: Study 1 Results: Mean of Performance Accuracy by Visual Aid (left), Mean of Performance Accuracy by Visual Aid in Each Dance (center), and Mean of Performance Accuracy by Dance (right)

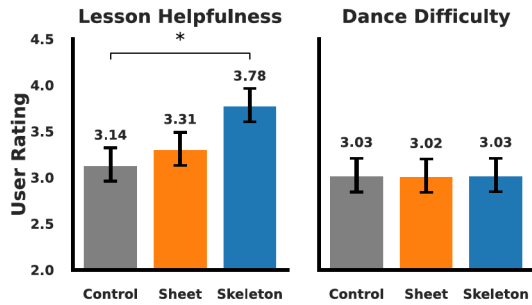


Figure 9: Study 1 Results: User-rated Lesson Helpfulness (left) and Dance Difficulty (right) estimated marginal means by Visual Aid

Helpfulness. To assess the effects of lesson type and dance on users' ratings of lesson helpfulness, we modeled treating lesson type and dance as fixed effects and including random effects for individual subject intercepts. The assumptions of homoscedasticity and normality were confirmed via Levene's test and residual plot inspection ($F_{8,121} = 1.163$, $p = 0.330$). The model's conditional R^2 was 0.171, with an AIC of 400.428, indicating moderate explanatory power and prediction accuracy. Estimated marginal means, illustrated in Fig. 9 (left), showed auto-generated lessons with skeleton aid were rated significantly more helpful than TikTok tutorials (Estimate = 0.65, $p = 0.009$), approximately 0.65 points higher on a five-point Likert scale. No significant difference was found between sheet motion aid lessons and TikTok tutorials, nor any significant effects of dance or interaction between dance and lesson type. Skeleton-overlay lessons were perceived as more helpful, while sheet motion aid lessons did not differ significantly from TikTok tutorials.

Difficulty. To assess the effects of lesson type and dance on perceived dance difficulty, we again modeled dance and lesson type as fixed effects and intercepts for each individual subject as random effects. Homoscedasticity and normality assumptions were confirmed by Levene's test and residual analysis ($F_{8,121} = 0.735$,

$p = 0.661$). The model demonstrated substantial explanatory power ($R^2 = 0.431$) and prediction accuracy (AIC=399.099). As shown in Fig. 9 (right), there was no significant main effect of lesson type on difficulty ratings, confirming that the counterbalanced design effectively separated perceived lesson helpfulness from perceived dance difficulty. Difficulty ratings did vary across dances, however, the only statistically significant comparison was dance C rated more difficult than dance A (Estimate=0.85, $p = 0.033$), which is convergent with our complexity metrics (Table 3). No interaction between dance and lesson type was detected, indicating that no dance was rated more difficult when paired with a given lesson type compared to the others.

Qualitative Feedback. The frequency of the codes in the users' responses in each condition highlighted several beneficial features of our auto-generated lessons, as shown in Fig. 10. For the question "What did you find helpful about this learning experience?", users particularly valued the slow-speed practice and repetition feature, which was available in all lesson types. Positive feedback on slow-motion practice appeared the most frequently across the lesson types. Some users explicitly highlighted the benefit of gradual speed increases during their practice sessions, saying, "The first time when it went from 50 to 100, I was surprised at how fast the song was. The 50, 75, and then 100 was extremely helpful in bringing the moves up to speed."

In the auto-generated lessons, the segmentation feature received praise from more than 16 users in both versions. One user commented, "I found that it was broken down into very small increments, which was very helpful in learning the faster-paced dance." Another participant appreciated the integration of segment learning into whole-dance learning, saying "The segments and slow speed while learning were helpful since they built up the dance in a digestible way." Feedback on the two types of visual aids—skeleton overlay and sheet motion—sharply differed. Eighteen users valued the skeleton feature as a clear tool for understanding dance movements. One user emphasized, "I enjoyed the stick figures because they took out the distractions from the video and I can see myself." However, sheet motion did not receive any positive feedback from users.

When users were asked what could be improved in the control lesson, eighteen participants expressed the need for segmentation

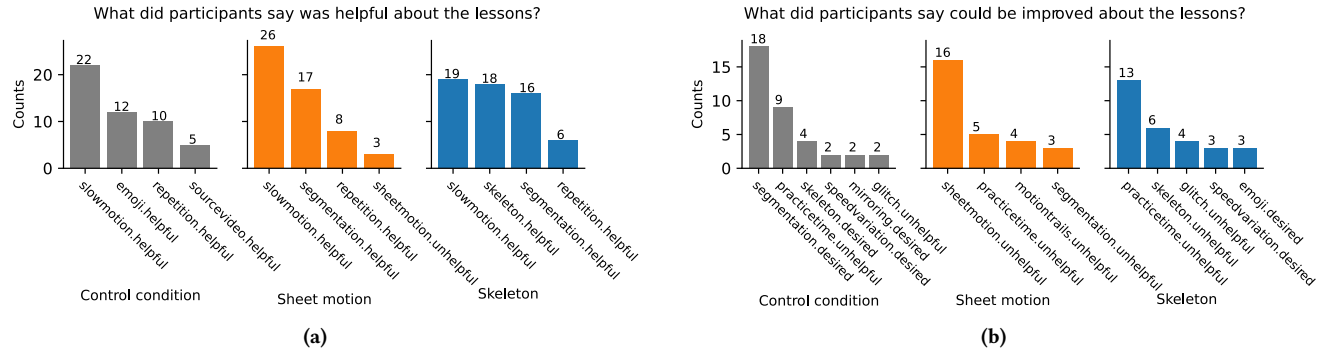


Figure 10: Study 1 Qualitative Results

in their learning as shown by the code `segmentation.liked`. By referring to “modules”, “components”, “chunking”, “sections”, and “step-by-step”, users communicated a preference for segment learning in the auto-generated lessons over the TikTok tutorial lesson. In the sheet motion aid lesson, a majority of users (16 participants) suggested removing the sheet motion aid. Users reported difficulty understanding the motion from sequences of static images and found the hand-motion trails distracting. One user compared the sheet motion aid with the skeleton aid, saying, “I thought that the practice with images was for the most part unhelpful because they were static and hard to reference. I much preferred the stick figure aid from the first dance.” Although the majority of users found the skeleton aid useful, six users expressed negative feelings about the skeleton aid when asked what could be improved in the lesson. One user explained, “The skeletons were useful, but they made it difficult to perceive depth, particularly when hip motions went forwards or backwards, so I felt like I didn’t learn that part very well.”

Twelve users found the emoji labels in the control lesson helpful. One commented, “I like the emojis for the dance. That helped me remember what to do next while learning it.” A similar positive perception emerged when the emojis were unavailable in the auto-generated lessons, with five users reporting that they wished emoji labels were included in the other two conditions.

5.3 Interpretation

In study 1, we designed and implemented a literature-informed dance learning system and examined its effectiveness with 52 users. Participants’ performance, evaluated by our automatic rating system, showed that the auto-generated lessons significantly enhanced learning outcomes compared to unaltered TikTok tutorials. Notably, in dances A and C, performance ratings were significantly higher with auto-generated lessons, indicating improved accuracy and retention even with a short learning period. These findings highlight the system’s ability to enhance motor skill acquisition through segmentation, adaptive guidance, and visual cues, as supported by prior research [1, 11, 17, 34]. The system’s automatic generation and presentation of practice plans further demonstrate its potential to scale the educational value of dance videos. Additionally, the system’s techniques for segmentation, lesson compilation, and presentation are adaptable to other types of motion videos, broadening its applicability.

We also sought to investigate user’s perception of learning experience in auto-generated lessons and TikTok tutorial videos. Users’ responses to the open-ended questions revealed that they identified slow speed and repetition as the main features they found most beneficial in across the control and auto-generated lessons—see Fig. 10, inline with prior multimedia learning [22] and motor learning [11] findings. Comparing the two types of auto-generated lessons, users achieved higher similarity scores and rated the learning experience as more helpful with the skeleton aid compared to TikTok tutorial videos, but not so with the sheet motion aid.

Responses to open-ended questions revealed that users perceived the skeleton visual aid to be helpful whereas the sheet motion aid was perceived as confusing and distracting.

Our results indicate that the effect of visual aids in auto-generated lesson on performance rating was impacted by the dance being learned. For dance A, users performed better with auto-generated lesson with sheet motion aid than with the TikTok tutorial, while for dance C, they performed better with auto-generated lesson with skeleton aid than sheet motion aid and TikTok tutorial. This suggests that these two types of visual aid provide distinct forms of support whose helpfulness varies depending on the dance. Skeleton overlay is a form of implicit-concurrent feedback that is intended to help during the initial stages of learning a dance. It provides an intermediary scaffolding between following a demo video and dancing without an aid—providing less information that the full video while still cuing the moves and timing—and seems to be effective in this role with our current implementation especially with dance C, the one which was perceived as the most difficult, and the one with the highest complexity (Table 3). This aligns with [37], which showed that concurrent feedback can be especially useful for complex tasks. We theorize that the skeleton overlay may help users by making the essential information more accessible, thus decreasing cognitive overload in a complex task and providing a stepping stone towards automaticity.

In contrast, sheet motion presents keyframes with motion trail cues, showing all keyframes simultaneously. In this way, compared to a skeleton-overlay or simple demonstration video, sheet motion simultaneously shows less information about the current target pose while showing more information overall. We think that this was likely an overwhelming visual display for a learner to follow along to. This is supported by users’ comments in the open-ended questions, where they referred to sheet motion as ‘static’, ‘frozen’,

Table 5: Study 2 Conditions: A 2x2 Factorial Design

		Segmentation	
		No	Yes
Labeling	No	Control	Segmentation-Only
	Yes	Emoji-Only	Segmentation-Emoji

and causing ‘sensory overload.’ Given dance A’s slower and less complicated nature (Table 3), users may have been able to follow along to the sheet motion despite the visual overload, thus productively attempting the dance with increased challenge. This is particularly interesting as users achieved significantly better performance with sheet motion in dance A despite perceiving the aid as unhelpful. Such a discrepancy is reported in other studies [6, 16], especially among inexperienced learners [30]. Overall, the evidence does not support the helpfulness of sheet motion as used by our current system.

Emoji labels in the control tutorial videos emerged as a key theme in the qualitative feedback. While these labels were retained to create an ecologically valid control condition, study 1 did not investigate whether users’ positive perceptions of emoji-labels corresponded to measurable learning benefits. This raises an important question: *how do emoji segment-labels influence learning, and are they linked to improved outcomes?* Answering this question is particularly relevant for automatic teaching systems, as evidence supporting the effectiveness of symbolic segment labels could justify developing methods to automatically generate and incorporate such labels from motion data.

6 Study 2: Investigation of Features of TikTok Dance Tutorials

Inspired by the unaddressed questions arising from participants’ qualitative responses in study 1, study 2 shifts focus to exploring specific features of TikTok dance tutorials—namely, emoji segment labels and dance segmentation—and their effect on learning outcomes. As such, we use a customized practice plan intended to mimic the experience of learning from these tutorial videos.

Motor learning theory suggests that segmenting complex tasks into smaller, manageable units can significantly enhance the learning process by allowing learners to focus on mastering one segment at a time before progressing [11]. In the context of TikTok dance challenge tutorial videos, the segmentation implied by overlaid emoji labels naturally supports this theory by dividing dances into smaller units, providing clear, focused intervals for practice and repetition. Such segmentation is predicted to aid in the gradual acquisition of complex movements, allowing learners to build proficiency incrementally and with greater precision. While segmentation in TikTok tutorial videos potentially enhances learning by breaking down complex dance moves into manageable units, it’s unclear if their inclusion of time-synchronized, color-coded emoji labels supports this learning process effectively.

On one hand, establishing dual codes to motions (through verbal cues, visual imagery, or sounds) has been shown improve memorization of motor tasks [7, 33]. Emojis could act as a dual code to

enable easier acquisition and retention of new moves in a choreography. Yet, emojis may be processed differently than other forms of dual coding studied in the literature – in particular, Homann et al. [12] evidence for both visio-spatial and verbal processing of emojis. Theories on multimedia learning describe separate processing mechanisms for visual and language-like information, and recommend a learning strategy that doesn’t overload either mechanism [22]. Given the visual processing demands of learning from observation it’s possible that simultaneous presentation of emojis as a visual cue could distract from the main learning task and impede learning.

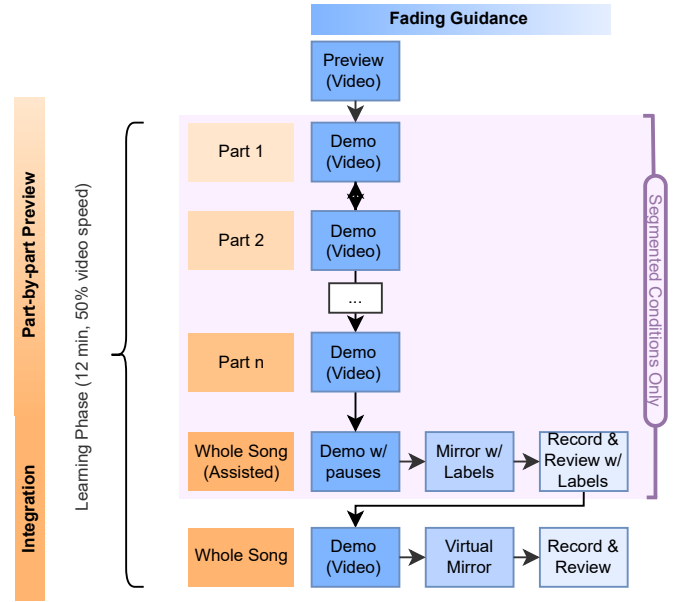


Figure 11: Practice Plan for Study 2. The structure of the practice plan uses fading feedback to support learning of the dance as a whole, but does not incorporate fading guidance in the segment-by-segment activities. The structure is intended to replicate the experience of learning from TikTok video tutorials, while experimenting with the segmentation and emoji-label features.

6.1 Practice Plan

The practice plan for study 2 starts with a preview activity, to give participants context as to the dance they’re learning. For segmented conditions, this was followed by a practice activity for each of the dance segments. The timing and duration of these segments was copied from the reference tutorial videos, representing the TikTok creators’ chosen segmentation.

The participants completed the study with the same procedure as the user study 1, learning a dance under each of four conditions: control condition, emoji-only condition, segmentation-only condition, and combined condition, per Table 5. In the control condition, users were presented with a version of the Tiktok tutorial videos with the emoji labels obscured and a single practice activity that played the entire dance. In the emoji condition, users were also

presented with a single practice activity, but the emoji-labels were left visible. In the segmented conditions, users were presented with activities for learning each segment individually, an integrative-practice activity that paused at segment-boundaries, and then the same practice activity as the unsegmented conditions. In the emoji-and-segmentation condition, the segments were labeled with the emoji from the tutorial videos, while in the segmentation-only condition, the segments were labeled with ‘Part 1, Part 2, ...’ labels. Users were given 12 minutes to learn the dances, and all learning activities and the recorded uploads were performed at half speed of the original dances. After each trial, participants were prompted to rate the helpfulness of the system and difficulty of the dance on a 10-point scale and answer the same open-ended questions as study 1.

6.2 Results

We fitted a linear mixed-effects model (LMM) to assess whether emoji labeling and dance segmentation significantly predicted participants’ performance accuracy as judged by the automatic similarity scores. The model included emoji labeling, segmentation, and dance as fixed effects, along with their interactions, with participants included as random intercepts.

Emoji labeling and segmentation were dummy-coded as binary factors, while the four dances (A, B, C, and D) employed deviation coding. This structure allows us to evaluate the mean accuracy score of each lesson directly compared to the mean across all dances. Levene’s test and residual plot inspection confirmed the model’s compliance with normality and homogeneity of variance assumptions ($F_{15,131} = 0.864$, $p = 0.605$). The model demonstrated moderate explanatory power (conditional $R^2 = .578$). This model was employed to calculate estimated marginal means, predicting participant performance accuracy in each of possible combination in the factorial design and dances. Fixed effects analysis showed a significant main effect for dance ($X^2(3, N = 147) = 0.9.227$, $p = 0.026$), suggesting variance in performance across dances—see Fig. 12 (right), yet found no significant effects for emoji labeling, segmentation, or their interactions—see Fig. 12 (left and center). This indicates participants’ performance was consistent across emoji and segmentation aids, with variations observed only across different dances. A Tukey HSD test was conducted to identify performance differences between dances. Findings indicated significantly lower performance in dance C versus dance B—see Fig. 12 (right). Considering dance C’s complexity and user-rated difficulty, performance differences highlight its challenging nature. Notably, significant performance disparity existed solely between dances C and B. Given similar scores for dances B, A, and D, it seems a performance plateau was reached, with dance C posing a unique challenge.

Pairwise comparisons between emoji labeling and segmentation within each dance were conducted using a Tukey HSD test to determine if there were differences in users’ performance attributable to either factor (see Fig. 12, center). The results showed no significant performance differences in any of the pairwise comparisons within each dance.

Helpfulness. To assess the impact of lesson type and dance on users’ perceptions of helpfulness, we again employed a model with these factors as fixed effects and individual subjects’ intercepts as

random effects. Levene’s test and a visual check of the residual plots confirmed our model did not violate the assumptions for homoscedasticity or normality ($F_{15,131} = 1.001$, $p = 0.454$). The model demonstrated a conditional R^2 of 0.542 and an AIC value of 540.346, indicating a moderate fit. Estimated marginal means for each lesson type are presented in Fig. 13 (left). No significant main effects of lesson type or dance, nor interactions between them, were found in predicting helpfulness ratings. This result indicates that users perceived all types of lessons as equally helpful regardless of the presence of emoji visual aids or segmentation.

Difficulty. To predict user ratings of dance difficulty, dance and lesson type were modeled as fixed effects, with individual subjects’ intercepts as random effects. Levene’s test and visual inspection of the residual plots did not reveal any violations of the assumptions of homoscedasticity or normality ($F_{15,131} = 0.733$, $p = 0.746$). This model exhibited a conditional R^2 of 0.563 and AIC of 536.515. The estimated marginal means of each type of lesson was shown in Fig. 13 (right). There was a main effect of dance on the difficulty ratings, meaning that the dances were not perceived as equally difficult. Consistent with the findings of study 1, this effect was driven by the contrast between dance C and A (Estimate = -1.86 , $p < 0.001$). There was no significant main effect of lesson type observed in this model. In other words, users perceived that the dance difficulty was not different across the lesson types. This was expected given our counterbalanced design. There was also no significant interaction between dance and lesson type, suggesting that no particular pairing of lesson type and dance yielded different perceived difficulty as compared to other possible combinations.

Qualitative Feedback. The analysis of users’ responses across the four tested conditions (visualized in Fig. 14) revealed important insights into how participants perceived the lessons. Interestingly, user preferences sometimes conflicted with their actual performance outcomes.

When asked what was helpful in each lesson, users consistently appreciated the slow-speed practice across all four lesson types. This feature was especially appreciated in the control lesson, with 18 participants mentioning it. Although there was no significant evidence that emojis or segmentation aids directly influenced subjective helpfulness ratings, many users described these features as beneficial in their qualitative responses. The emoji feature, in particular, was frequently praised when included in the lesson. Users found emojis helpful for visualizing and remembering dance moves. One participant noted, “The varieties of emojis helped me to remember the difficult steps and to perform correctly,” while another said, “The emojis helped very much. It gave the direction to take the steps. It was very simple and nice.” Additionally, four participants expressed a desire for accompanying text instructions to further clarify dance movements. Some suggested combining text with segmentation, commenting, “Could use extra written/verbal instructions about each step,” or “voice instruction on steps.”

Most users recognized segmentation as advantageous in both the segmented and combination lessons. They particularly appreciated practicing dances in manageable segments, which helped prevent feelings of being overwhelmed. One user noted, “I loved how the dance was split up into easy-to-digest parts.” Segmented practice

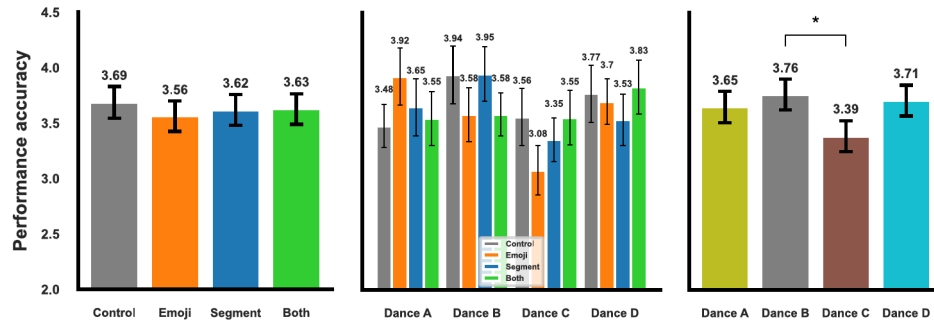


Figure 12: Study 2 Results: Mean of Performance Accuracy by Visual Aid (left), Mean of Performance Accuracy by Visual Aid in Each Dance (center), and Mean of Performance Accuracy by Dance (right)

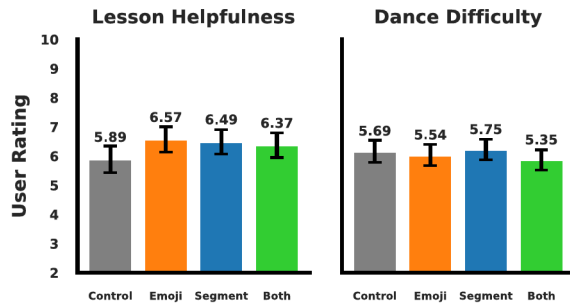


Figure 13: Study 2 Results: User-rated Lesson Helpfulness (left) and Dance Difficulty (right) estimated marginal means by Visual Aid

was also seen as useful to focusing on specific movements that required extra practice. One commented, “I liked how if I had trouble with certain parts I could look back at certain steps.” Additionally, many users valued the inclusion of pauses between segments, especially when used in conjunction with segmentation. One participant emphasized, “I like how the tutorial broke the dance into steps and paused between each step because it allowed me to process the dance,” highlighting how the pauses helped them connect individually learned segments.

When asked what could be improved in each lesson, participants recommended adding segmentation or emojis to lessons that lacked these features, demonstrating their perceived value in the learning process. In the control lesson, segmentation and emojis were the most commonly mentioned areas for improvement, suggesting a strong user preference for these aids. One participant explained, “I need to learn ONE thing at a time! So I wanted to learn the first move, then the second, then the third, then... put it all together at the end. It was way too hard to learn it all at once and not be able to pause the video.” Similarly, participants recommended adding emojis or segmentation in conditions where one of these features was missing. However, some users expressed mixed feelings about segmentation, suggesting improvements for its implementation, as

reflected in the counts for the code `segmentation.refine`. They called for adjustments in segment size, with one participant noting, “I think some of the moves could be broken down further,” while another suggested, “Lesser segmentation.”

6.3 Interpretation

The central finding of user study 2 is that neither the emoji-annotation nor segmentation features of TikTok dance tutorial videos appear to enhance objective measures of dance learning, whether incorporated into the teaching system independently or together. Neither emojis nor segmentation were predictive of participants’ performance or helpfulness ratings. This result seems to be inconsistent with predictions of motor learning theory.

Rather than challenging established research in motor learning, we interpret the null findings in the second study as evidence that segmentation of a dance alone is insufficient to improve learning outcomes—this segmentation must also be meaningfully incorporated into a practice routine to enable part learning. As implemented in the study 2, the segment learning activities only contain a single video-following step (see Fig. 11), as opposed to the multi-step segment-learning design in study 1 (Fig. 7). This simplification was done intentionally in order to match the typical experience of observational learning from video, in which the aids and progression found in the study 1 practice plan would not be present. The null finding also serves as a cautionary note that illustrates how the complexities of a fully-realized system can nullify the theoretical intent of an assistive feature.

While participants found emojis helpful for remembering movements, they also suggested that supplementary textual instructions could improve clarity, highlighting the limitations of emojis in representing complex motions like dance. Due to their simplicity and limited variability, emojis may contribute to extraneous processing, particularly when combined with simultaneous activities such as sheet motion. These findings suggest that emojis might be more effective in a pretraining step, where learners observe the dance and its emoji representation to build a mental model before performing it.

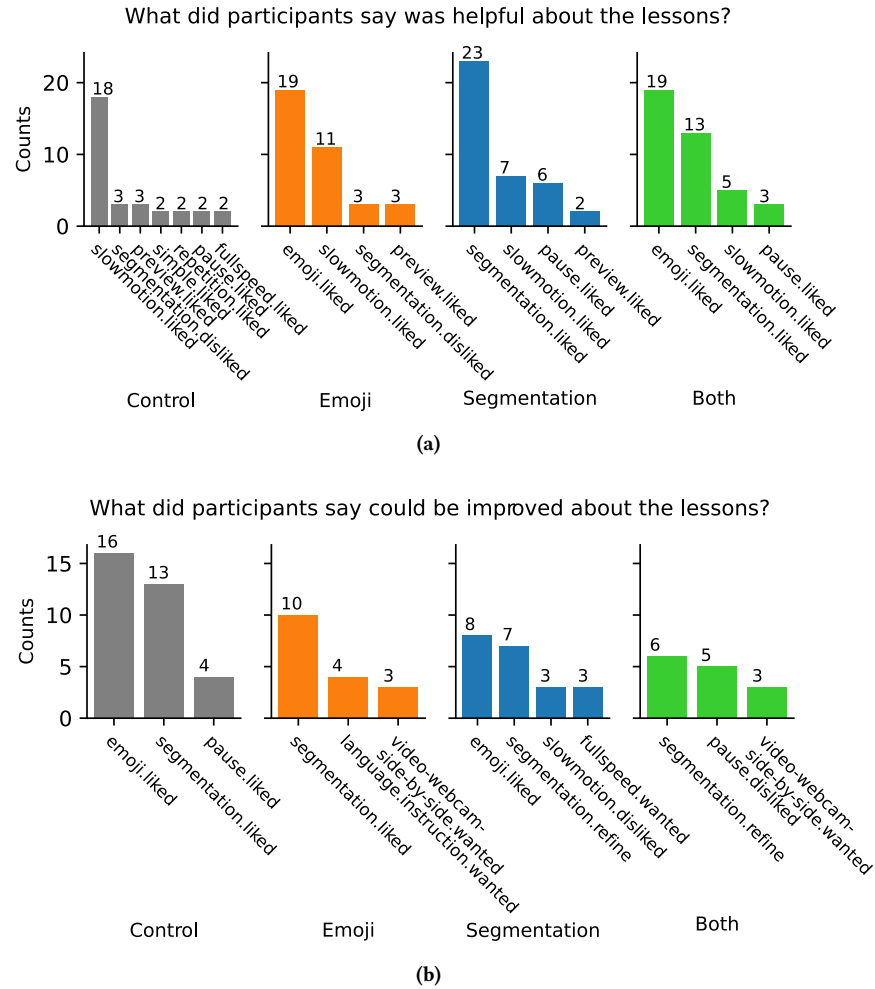


Figure 14: Study 2 Qualitative Results

7 Discussion

The process of designing, implementing, and evaluating this automatic dance teaching system has led us to insightful takeaways regarding the operationalization of learning experience design in the context of motor skill learning from video. Study 1 showed evidence that our system, with its assistive features, was perceived as more helpful and led to better user performances than standard video tutorials. However, when we examined two features (segmentation and emoji labels) common to TikTok tutorial videos in user study 2, we observed no such increase in performance—despite the users’ sentiments that these features were helpful. This suggests that while engaging, these learning aids may require more thoughtful integration to meaningfully impact learning outcomes.

7.1 Design Opportunities

Our findings highlight several actionable design opportunities for future movement learning systems.

Prioritize Part-Learning for Complex Movements. Our results indicate that segmentation and part-learning features are most effective

when integrated into a structured lesson plan that encourages users to incrementally master segments of a movement before progressing to the full sequence. This aligns with existing motor learning theories, which suggest that incremental part-learning is beneficial for tasks with high complexity. Incorporating substantial part-learning activities is essential for supporting learners in building up their proficiency over time.

Enhance Visual Aids While Managing Cognitive Load. Study 2 showed that although users found visual aids like emoji labels helpful, they did not significantly enhance performance. This suggests that these aids must be carefully integrated within a well-structured lesson plan to be effective. To optimize the effectiveness of visual aids, designers should consider introducing these aids in stages—perhaps as part of pre-rehearsal activities—while ensuring that they do not oversaturate learners’ visual input streams during practice.

Offer User-Controlled Segmentation and Playback Speed. Given the variability in learners’ preferences for segmentation granularity and playback speed, future systems should allow users to adjust these parameters to fit their individual learning styles.

Utilize Simplified Visual Cues for Key Movement Phases. Users expressed confusion when presented with overly complex visual representations, such as sheet motion displays. Simplifying visual representations to focus on key poses or movement phases may improve learners' ability to follow along with dynamic movements.

Include a Pre-Rehearsal Step for Familiarization. We recommend incorporating a pre-rehearsal step that familiarizes learners with key labels or segments before engaging in live movement practice. Qualitative feedback indicated that emoji labels were appreciated, but their effectiveness was limited without prior exposure. A pre-rehearsal step could provide an opportunity to introduce these semantic cues early on.

7.2 Limitations & Future Research

TikTok dances, predominantly short and stylistically narrow, may not fully represent the extensive learning curves required for more complex dance forms. However, many TikTok dance tutorials share common characteristics, such as distinct musical phrasing, segmented choreography, and repetitive movement patterns. For these types of dances, our findings on segmentation, visual aids, and structured practice plans are likely to be applicable. Nonetheless, there remains an opportunity to explore longer, more challenging dance content. Additionally, considering TikTok's younger demographic [9] and our study's reliance on university students, future research should broaden participant diversity to include a wider range of ages and backgrounds.

Future work might consider the following research questions: (1) *How do different demographics engage with and benefit from video-based motor learning, and what adaptations are necessary to meet diverse needs?* (2) *Can extended, more complex dance sequences enhance long-term skill retention and mastery compared to shorter, social media-style dances?* (3) *In what ways can video-based learning systems be optimized for other forms of motion, and what are the key factors in designing content-specific instructional aids?*

The limited scope of Study 2's findings, particularly concerning the specific context and application of TikTok dance challenge videos, highlights areas for broader investigation. In our experimental setup, designed to reflect typical video learning environments, neither emoji-labeling nor segmentation features significantly improved performance. However, these features might still offer benefits if applied differently. Study 1 illustrates this, showing enhanced outcomes with practice plans that combine segmentation and visual aids.

Future studies should explore various approaches to implementing labeling and segmentation to determine their impact more precisely. Worthwhile research questions for future work may include: (4) *how do people approach learning from social media dance videos in naturalistic settings? What technologies and practice strategies are adopted?* (5) *can segment or movement labels enhance motor learning when incorporated into a pre-training task? What are the necessary conditions for dual-coding of motions to be effective?* (6) *are user defined segmentations and segment labels preferable to system-assigned ones in the learning process?*

7.3 Conclusion

Broadly, these two studies reflect the versatility of our system when it comes to investigating learning science questions of dance learning. By tailoring the practice plan to specific study objectives, we were able to investigate substantially different empirical questions with minimal system modifications. The system's automatic performance scores, movement complexity quantification, and LLM-based feedback summarization facilitated efficient, scalable comparisons across conditions. The high degree of correlation between the automatic performance rating and human similarity judgments is a noteworthy finding, validating the metric as a reliable alternative to laborious manual ratings. This real-time computation capability provides automated dance practice systems with a validated, practicable means to quantify the similarity of learners' performances to reference dances. Together, these findings emphasize the importance of aligning assistive features with learners' attentional capacities and pedagogical goals, underscoring the potential of our system to advance research and practice in dance motor learning. By refining these tools and expanding their application to other movement domains, we can continue to advance both the study of motor learning and the design of scalable, video-based educational systems for movement.

Acknowledgments

This work was funded by the National Science Foundation under grant NSF 2202553 to DJMK, XZ, and DB. The authors also give thanks to L. Hennessy, H. Jhaveri, T. Brent, B. Aubrey, S. Young, and K. Romanger for help with data collection and video ratings.

References

- [1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) (UIST '13). Association for Computing Machinery, New York, NY, USA, 311–320. doi:10.1145/2501988.2502045
- [2] J.B. Blanchet and Sixuan Han. 2023. Integrating a LLM into an Automatic Dance Practice Support System: Breathing Life Into The Virtual Coach. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 118, 2 pages. doi:10.1145/3586182.3625119
- [3] J.B. Blanchet, Megan Hillis, Yeongji Lee, Qijia Shao, Xia Zhou, David Kraemer, and Devin Balkcom. 2023. Automatic Generation and Teaching of Dance Lessons from Video. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications* (Newport Beach, California) (HotMobile '23). Association for Computing Machinery, New York, NY, USA, 145. doi:10.1145/3572864.3581592
- [4] J.B. Blanchet, Megan E. Hillis, Yeongji Lee, Qijia Shao, Xia Zhou, David J. M. Kraemer, and Devin Balkcom. 2023. LearnThatDance: Augmenting TikTok Dance Challenge Videos with an Interactive Practice Support System Powered by Automatically Generated Lesson Plans. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (<conf-loc>, <city>San Francisco</city>, <state>CA</state>, <country>USA</country>, </conf-loc>) (UIST '23 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 84, 4 pages. doi:10.1145/3586182.3615801
- [5] John Brooke. 1996. Sus: a 'quick and dirty' usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.
- [6] Christine E Carter and Jessica A Grahn. 2016. Optimizing music learning: Exploring how blocked and interleaved practice schedules affect advanced performance. *Frontiers in psychology* 7 (2016), 1251.
- [7] James M Clark and Allan Paivio. 1991. Dual coding theory and education. *Educational psychology review* 3 (1991), 149–210. doi:10.1007/BF01320076
- [8] Christopher Clarke, Doga Cavdir, Patrick Chiu, Laurent Denoue, and Don Kimber. 2020. Reactive video: adaptive video playback based on user motion for supporting physical activity. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 196–208.

- [9] Brian Dean. 2024. *TikTok Statistics You Need to Know*. <https://backlinko.com/tiktok-users#key-tik-tok-user-demographics>
- [10] Paul M Fitts and Michael I Posner. 1967. Human performance. (1967).
- [11] Fabio E Fontana, Ovide Furtado Jr, Oldemar Mazzardo, and Jere D Gallagher. 2009. Whole and part practice: A meta-analysis. *Perceptual and motor skills* 109, 2 (2009), 517–530. doi:10.2466/PMS.109.2.517-530
- [12] Lauren A Homann, Brady RT Roberts, Sara Ahmed, and Myra A Fernandes. 2022. Are emojis processed visuo-spatially or verbally? Evidence for dual codes. *Visual Cognition* 30, 4 (2022), 267–279. doi:10.1080/13506285.2022.2050871
- [13] Yueqi Hu, Shuangyuan Wu, Shihong Xia, Jinghua Fu, and Wei Chen. 2010. Motion track: Visualizing variations of human motion data. In *2010 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 153–160.
- [14] Jun Inagaki, Toshihiro Nakajima, Hirokazu Haruna, Keisuke Kon, Yoichiro Sato, and Sadayuki Hongo. 2019. A Prototype of a Motor-Learning Support System Using a Motion-Capture Device. In *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*. 1148–1150. doi:10.1109/GCCE46687.2019.9015368
- [15] Alexandros Kitsikidis, Kosmas Dimitropoulos, Deniz Uğurca, Can Bayçay, Erdal Yilmaz, Filareti Tsalakanidou, Stella Douka, and Nikos Grammalidis. 2015. A game-like application for dance learning using a natural human computer interface. *Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being* (2015), 472–482. doi:10.1007/978-3-319-20684-4_46
- [16] Nate Kornell and Robert A. Bjork. 2008. Learning Concepts and Categories: Is Spacing the “Enemy of Induction”? *Psychological Science* 19, 6 (2008), 585–592. doi:10.1111/j.1467-9280.2008.02127.x arXiv:https://doi.org/10.1111/j.1467-9280.2008.02127.x PMID: 18578849
- [17] Matthew Kyan, Guoyu Sun, Haiyan Li, Ling Zhong, Paisarn Muneesawang, Nan Dong, Bruce Elder, and Ling Guan. 2015. An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 2 (2015), 1–37.
- [18] William Li, Lyn Bartram, and Philippe Pasquier. 2016. Techniques and Approaches in Static Visualization of Motion Capture Data. In *Proceedings of the 3rd International Symposium on Movement and Computing* (Thessaloniki, GA, Greece) (MOCO '16). Association for Computing Machinery, New York, NY, USA, Article 14, 8 pages. doi:10.1145/2948910.2948935
- [19] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019). doi:10.48550/arXiv.1906.08172
- [20] Richard Magill and David I Anderson. 2010. *Motor Learning and Control: Concepts and Applications*. McGraw-Hill Publishing New York.
- [21] Zoe Marquardt, João Beira, Natalia Em, Isabel Paiva, and Sebastian Kox. 2012. Super Mirror: A Kinect Interface for Ballet Dancers. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI EA '12). Association for Computing Machinery, New York, NY, USA, 1619–1624. doi:10.1145/2212776.2223682
- [22] Richard E Mayer. 2017. Using multimedia for e-learning. *Journal of computer assisted learning* 33, 5 (2017), 403–423.
- [23] Luis Molina-Tanco, Carmen García-Berdónes, and Arcadio Reyes-Lecuona. 2017. The Delay Mirror: A Technological Innovation Specific to the Dance Studio. In *Proceedings of the 4th International Conference on Movement Computing* (London, United Kingdom) (MOCO '17). Association for Computing Machinery, New York, NY, USA, Article 9, 6 pages. doi:10.1145/3077981.3078033
- [24] Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 152–158.
- [25] Katerina El Raheb, Marina Stergiou, Akrivi Katifori, and Yannis Ioannidis. 2019. Dance Interactive Learning Systems: A Study on Interaction Workflow and Teaching Approaches. *ACM Comput. Surv.* 52, 3, Article 50 (6 2019), 37 pages. doi:10.1145/3323335
- [26] Katerina El Raheb, George Tsampounaris, Akrivi Katifori, and Yannis Ioannidis. 2018. Choreomorphy: A whole-body interaction experience for dance improvisation and visual experimentation. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. 1–9.
- [27] Rémi Rampin and Vicky Rampin. 2021. Taguette: open-source qualitative data analysis. *Journal of Open Source Software* 6, 68 (2021), 3522. doi:10.21105/joss.03522
- [28] Jean-Philippe Rivière, Sarah Fdili Alaoui, Baptiste Caramiaux, and Wendy E. Mackay. 2018. How Do Dancers Learn To Dance? A First-Person Perspective of Dance Acquisition by Expert Contemporary Dancers. In *Proceedings of the 5th International Conference on Movement and Computing* (Genoa, Italy) (MOCO '18). Association for Computing Machinery, New York, NY, USA, Article 6, 7 pages. doi:10.1145/3212721.3212723
- [29] Jean-Philippe Rivière, Sarah Fdili Alaoui, Baptiste Caramiaux, and Wendy E Mackay. 2019. Capturing movement decomposition to support learning and teaching in contemporary dance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–22.
- [30] Jean-Philippe Rivière, Sarah Fdili Alaoui, Baptiste Caramiaux, and Wendy E Mackay. 2019. Capturing movement decomposition to support learning and teaching in contemporary dance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–22.
- [31] Anna Rizzo, Katerina El Raheb, Sarah Whatley, Rosemary E. Kostic Cisneros, Massimiliano Zanoni, Antonio Camurri, Vladimir Viro, Jean marc Matos, Stefano Piana, Michele Buccoli, Amalia Markatzi, Pablo Palacio, Oshri Even-Zohar, Augusto Sarti, Yannis Ioannidis, and Edwin Morley-Fletcher. 2018. WhoLoDancE: Whole-body Interaction Learning for Dance Education. *CEUR Workshop Proceedings* (3 Nov. 2018), 41–50.
- [32] Sayed Kavos Salehi, Farshid Tahmasebi, and Fateme Sadat Talebrokni. 2021. A different look at featured motor learning models: comparison exam of Gallahue's, Fitts and Posner's and Ann Gentile's motor learning models. *Movement & Sport Sciences* 2 (2021), 53–63. doi:10.1051/sm/2021012
- [33] R.A. Schmidt, T.D. Lee, C. Winstein, G. Wulf, and H.N. Zelaznik. 2018. *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics. <https://books.google.com/books?id=EvJ6DwAAQBAJ>
- [34] Alessandra Semeraro and Laia Turmo Vidal. 2022. Visualizing instructions for physical training: Exploring visual cues to support movement learning from instructional videos. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [35] Simon Senecal, Niels A Nijdam, Andreas Aristidou, and Nadia Magnenat-Thalmann. 2020. Salsa dance learning evaluation and motion analysis in gamified virtual reality environment. *Multimedia Tools and Applications* 79 (2020), 24621–24643. doi:10.1007/s11042-020-09192-y
- [36] Qijia Shao, Amy Sniffen, Julien Blanchet, Megan E. Hillis, Xinyu Shi, Themistoklis K. Haris, Jason Liu, Jason Lamberton, Melissa Malzkuhn, Lorna C. Quandt, James Mahoney, David J. M. Kraemer, Xia Zhou, and Devin Balkcom. 2020. Teaching American Sign Language in Mixed Reality. 4, 4, Article 152 (Dec. 2020), 27 pages. doi:10.1145/3432211
- [37] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review* 20 (2013), 21–53. doi:10.3758/s13423-012-0333-8
- [38] Vikash Singh, Celine Latulipe, Erin Carroll, and Danielle Lottridge. 2011. The choreographer's notebook: a video annotation system for dancers and choreographers. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*. 197–206.
- [39] Il Hong Suh, Sang Hyoung Lee, Nam Jun Cho, and Woo Young Kwon. 2017. Measuring motion significance and motion complexity. *Information Sciences* 388–389 (2017), 84–98. doi:10.1016/j.ins.2017.01.027
- [40] Ramin Tadayon, Troy McDaniel, and Sethuraman Panchanathan. 2017. A Survey of Multimodal Systems and Techniques for Motor Learning. *Journal of Information Processing Systems* 13, 1 (2017), 8–25. doi:10.3745/JIPS.02.0051
- [41] Milka Trajkova and Francesco Cafaro. 2016. E-Ballet: Designing for Remote Ballet Learning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) (UbiComp '16). Association for Computing Machinery, New York, NY, USA, 213–216. doi:10.1145/2968219.2971442
- [42] Milka Trajkova and Francesco Cafaro. 2018. Takes Tutu to Ballet: Designing Visual and Verbal Feedback for Augmented Mirrors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 38 (mar 2018), 30 pages. doi:10.1145/3191770
- [43] Shuhei Tsuchida, Haomin Mao, Hideaki Okamoto, Yuma Suzuki, Rintaro Kanada, Takayuki Hori, Tsutomu Terada, and Masahiko Tsukamoto. 2022. Dance Practice System that Shows What You Would Look Like if You Could Master the Dance. In *Proceedings of the 8th International Conference on Movement and Computing*. 1–8.
- [44] Brady T West, Kathleen B Welch, and Andrzej T Galecki. 2022. *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC. doi:10.1201/9781003181064
- [45] Gabriele Wulf and Rebecca Lewthwaite. 2016. Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic bulletin & review* 23 (2016), 1382–1414.
- [46] Gabriele Wulf, Charles Shea, and Rebecca Lewthwaite. 2010. Motor skill learning and performance: a review of influential factors. *Medical education* 44, 1 (2010), 75–84. doi:10.1111/j.1365-2923.2009.03421.x
- [47] Yang Yang, Howard Leung, Lihua Yue, and Liqun Deng. 2010. Evaluating Human Motion Complexity Based on Un-Correlation and Non-smoothness. In *Advances in Multimedia Information Processing - PCM 2010*, Guoping Qiu, Kin Man Lam, Hitoshi Kiya, Xiang-Yang Xue, C.-C. Jay Kuo, and Michael S. Lew (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 538–548. doi:10.1007/978-3-642-15696-0_50
- [48] Qiushi Zhou, Cheng Cheng Chua, Jarrod Knibbe, Jorge Goncalves, and Eduardo Velloso. 2021. Dance and choreography in HCI: a two-decade retrospective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [49] Qiushi Zhou, Louise Grebel, Andrew Irlitti, Julie Ann Minaai, Jorge Goncalves, and Eduardo Velloso. 2023. Here and Now: Creating Improvisational Dance Movements with a Mixed Reality Mirror. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 183, 16 pages.

doi:10.1145/3544548.3580666

- [50] Zhongyi Zhou, Anran Xu, and Koji Yatani. 2021. SyncUp: Vision-based Practice Support for Synchronized Dancing. *Proceedings of the ACM on Interactive, Mobile,*

Wearable and Ubiquitous Technologies 5, 3 (2021), 1–25.