Thinking Through the Hands: An Exploratory Study of Hand Movements to Assess Students Problem-Solving in Mechanistic Reasoning Tasks

Harshil Safi*[1], Megha Bansal*[1], Madhu Vadali[1], Barbara Bruno[2], Aditi Kothiyal[1]

[1]Indian Institute of Technology Gandhinagar, India
[2]Karlsruhe Institute of Technology, Germany
harshil.safi@alumni.iitgn.ac.in, megha.bansal@iitgn.ac.in, madhu.vadali@iitgn.ac.in, barbara.bruno@kit.edu, aditi.kothiyal@iitgn.ac.in

Abstract

Theories of embodied learning emphasize that learning processes are grounded in bodily actions and interactions with the environment, suggesting that movements play a fundamental role in problem solving, decision making, and learning. This perspective holds particular relevance for making-based learning settings, where patterns of movement and spatial engagement can reveal strategic expertise. Prior research has examined distinctions between students who learned and did not learn, but manual coding of actions presents scalability and real-time application challenges. To address this gap, we develop a computer vision–based analysis pipeline for automated detection and characterization of hand movements during complex assembly tasks. In an exploratory study, we apply this approach to video data of students engaged in the assembly of a differential gearbox, quantifying metrics such as amount and speed of movement. Results indicate that learners show fewer right-hand movements than novices and exhibit reduced movement speed, with a progressive decline in speed as the task unfolds. Non-learners, by contrast, display more uneven hand movement speed. These findings, while preliminary, highlight measurable differences in actions of learners and non-learners, and therefore have potential implications for learning support. Specifically, the ability to computationally distinguish movement profiles can inform the design of adaptive learning interventions, providing real-time performance assessment and targeted feedback for making-based learning.

1 Introduction

Makerspaces are physical spaces where ideas are realised through prototyping, experimentation, mental and computer simulation, formal and informal knowledge integration, collaboration and reflection (Sinha and Chandrasekharan 2021). They are spaces where students have agency to work on problems of their interest, individually or collaboratively, and in the process acquire conceptual understanding (Sinha and Chandrasekharan 2021), skills such as creativity (Soomro et al. 2023) and problem-solving, (Sinha and Chandrasekharan 2021) and achieve a sense of belonging into engineering (Andrews, Borrego, and Boklage 2021). Thus, learning of concepts and skills by making artefacts or making-based learning is emerging as an important pedagogical approach at all levels of education (Schad and Jones 2020). However, the learning effectiveness of this approach depends upon learners receiving extensive and diverse kinds of support as they make, ranging from conceptual support and support for using the tools to affective support for persistence (Turakhia et al. 2024; Winters et al. 2023). Further, the effectiveness of making-based learning depends on learners' making behaviours – ie, the kinds of actions they do and the interactions they have with the environment and their collaborators (Davis et al. 2024; Worsley and Blikstein 2018). Recent theoretical perspectives characterising the mechanisms of making, suggest that making capabilities "can be understood as complex coagulations of internal and external processes related to basic actions, physical structures, and their affordances" (Chandrasekharan, Sinha, and Date 2025).

Research in making scenarios suggests that a higher proportion of certain groups of "implement" actions (which include undoing parts of a previous design or putting pieces together) are more productive for learning than other types of actions (Worsley and Blikstein 2018). Similar research also shows that certain kinds of actions (including meshing, rotating, mounting and making correct connections) underlie expertise in making tasks (Davis et al. 2024). Thus, we expect that different *types* of actions underpin different *levels* of making performance. However, both these studies rely partly on hand coding of students' actions, as automatic identification of making actions remains a challenge. On the other hand, research shows that automatically detected and processed audio-, video-, and tabletop log action-based features (such as quantity of speech) can be used to identify occurrences of behaviours productive or unproductive for learning (Nasir et al. 2021). Other research has shown that differences in the quantity of hand movement underlie differences in qualitatively different hand gestures during knowledge co-construction discourses (Lyu et al. 2025). Together, these findings suggest that even *low fidelity, automatically extractable features* like quantity of speech instead of quality of speech, and quantity of movement instead of type of gestures can be used to distinguish between different performance levels in learning scenarios.

In this work, we investigate this conjecture in the context of a making-based learning scenario. Leveraging computer vision models that can automatically detect human hand

*These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

landmarks, we present an exploratory investigation of the differences between learning and non-learning students, in their quantity and speed of movement. Our goal is to identify metrics which can be used to distinguish learners from non-learners, based on their quantity and speed of movement and intervene appropriately in making-based learning scenarios using automatic agents such as social robots.

2 Related Work

Hand movements and use of space in learning and educational environments has been of extensive interest to learning scientists, starting from qualitatively understanding how the use of space between the body and the screen among students shapes their exposure to resources, approach to the problem and their collaboration with peers (Davidsen and Christiansen 2014) to the analysis of problem solving strategies being determined by availability of gestures among undergraduates and adults (Alibali et al. 2011). Postures, body language, and smoothness of hand movements were found to be indicative of presentation skills of undergraduate students (Echeverría et al. 2014). Further, assuming open body postures was associated with higher scores in fluency, flexibility, and originality indexes of creative thinking (Andolfi, Di Nuzzo, and Antonietti 2017). Gestures and body movements, understood as dynamically produced body-material resources are effectively used for communication and illustration, as a cognitive and learning tool, as well as for instructing each other during collaborative tasks (Davidsen and Ryberg 2017). Gross application of gestures was in the form of numerical and spatial tools among children (Davidsen and Ryberg 2017). These studies show that the use of hands in space is indicative of both communication and cognition implicating its relevance in learning, which is a social-cognitive activity.

Qualitatively different gestures and their application were observed in high and low performing pairs of students in solving design problems (Lyu et al. 2025). The gestures were associated with different discourses in the task and strategies used by the pairs. Similarly, expert and novice approaches to problem solving implicate distinct actions and interactions with the objects of a task, which further predict performance (Davis et al. 2024). Specifically, for a novel assembly task, more instances of 'mechanical actions' and less instances of 'structural actions' differentiated experts from novices, while for a troubleshooting task, the more objects are interacted with, the more expert approach it implied. Gestures and their function is closely associated with the situation, tools and technology, and structure of the environment (Davidsen and Ryberg 2017). In prediction of gear movement, gestures promote and direct perceptual-motor strategies, and simulation of gears' physical movements and spatial positions (Alibali et al. 2011), and that activation of perceptual-motor information was increased with the production of gestures.

The increasing pedagogical emphasis on hands-on and making-based learning demands movement based measures of learning and progress. Popularly used standardized tests to evaluate skills of students are not useful measures in this case, whereas qualitative approaches using video-recorded sessions or student's written reports cannot be executed for large numbers of students or to monitor learning progress (Riquelme et al. 2019). For instance, in a multimodal learning analytics framework, various feedback analytics using spoken interventions and bodily movements of team members effectively demonstrate collaborative and competitive dynamics, dominant and passive engagements, and progress during the task for instructors and teachers (Noël et al. 2022).

While the fact that different gestures implicate different approaches in a problem-solving task has been researched (Alibali et al. 2011; Lyu et al. 2025), it is still an open question as to how the quantity of hand movement is associated with qualitatively different gestures and actions. This is a recent trend in research motivated by the need to have large scale applications of collaborative and making-based learning. In researching different types of collaborative problem-solving discourses, researchers found significantly larger amounts of hand movement in conflict-oriented discourses than quick, and integration-oriented ones (Lyu et al. 2023). Moreover, the ratio of hand movement within the pairs differed in the three discourses. The work was unique in its presentation of the feasibility of frame-by-frame analysis of body movement, here hands, to study engagement with the task. Machine-detected body movements as indicators of embodied mathematical reasoning and collaborative knowledge building were studied by computing the variances of bounding boxes of upper body joint movement (Sung and Nathan 2024). Larger variances were associated with the learning outcome to develop awareness about non-verbal cues and multimodal knowledge expression as indicators of mathematical knowledge. These analyses approach the quantification of hand movement by embedding them in evidence of different gestures and behaviours during learning and problem solving.

The work discussed above still warrants a manual contribution to data processing. To effectively use hand movements to predict learning outcomes, efforts to study these two quantitatively are required. Using a completely computational approach, Yao and Billard (2020) identified metrics related to differences in grasp, dexterity, positioning, and force applied and its angle in a watchmaking task, comparing the behaviours of novices and experts, and corroborated this through qualitative, visual inspection of selected moments in the task. In a similar fashion, Noël et al. (2022) presented a multimodal learning analytics platform that uses machine learning techniques on data recorded from multi-view cameras and multidirectional microphones to detect, recognize, and visualise information about speech and postures. The resulting visualizations were validated by teachers.

A comprehensive review of previous work shows that completely computational, computer vision based hand movement metrics provide scope for scaling active, in-task progress evaluations in making-based learning environments. However, the review also shows a lack of quantitative investigations of hand movements in learning environments and how they can be metrics of learning, and not just expertise or inter-personal behaviour. Therefore, in this pa-

per we explore whether quantities of hand movements can be indicative of learning. These computational approaches heavily depend on computer vision models to track the position and movement of hands. Some state of the art computer vision models are MediaPipe, HaMeR, FrankMoCap, and MobRecon (Niu et al. 2024; Lyu et al. 2023; Pavlakos et al. 2024). They are algorithms that segment and detect hand features (Oudah, Al-Naji, and Chahl 2020) using RGB or RGB-D data files (Carfi et al. 2018). In our work, we chose MediaPipe and our reasons are explained in subsection 3.6.

3 Methods

3.1 Research Question and Hypotheses

Aligned with our goal of developing fully computational markers for making-based learning, we investigate the following research question in this study: *Are there any differences between learners and non-learners in their total movement and speed of movement, over the entire duration and across different time segments, as they perform an making task?* This research question, based on literature such as (Yao and Billard 2020; Uemura et al. 2014) which suggest that experts show smoother movements, translates to the following hypotheses.

1. Learners' movements are slower than non-learners, i.e., lower speed of movement overall and in each segment and thus,
2. Learners show lower overall movement, and lower movement in each segment compared.

3.2 Making-Based Learning Task

In order to assess differences in movements, we assigned participants a task to assemble a small 3D printed differential gearbox from its parts, such as gears, shafts, and a housing, as shown in Figure 1. Each piece was embedded with magnets for easy assembly. The disassembled gearbox was presented in ten pieces, without any information about the final object, to authentically recreate a learning scenario encountered in a makerspace. The magnets allow many plausible but incorrect partial assemblies, while only one configuration produced a correctly assembled gearbox. During the assembly task, participants were instructed to not to pull apart the already attached pieces.

A differential gearbox divides input torque between two outputs and allows those outputs to rotate at different speeds. This task was chosen because it offers an authentic making activity that has, in past research, been shown to differentiate between experts and novices (Davis et al. 2024). Further, it aligns with concepts such as gear meshing, torque transfer and alignment, remains accessible to novices while informative for experts, and can be administered consistently with standardised materials and workspace.

3.3 Instruments

To test our hypothesis about making-based learning, we evaluated participants knowledge before (pre) and after (post) the assembly task using 10-item pen-paper tests. A set of two matched tests was prepared. The tests assess problem solving and conceptual knowledge embedded in the context of the experimental task. They were created by an expert Mechanical engineering teacher, consist of textbook-style conceptual questions (e.g., Norton 2009), were validated through a pilot study, and refined to address ceiling effects. Test administration was counterbalanced across participants. The tests had multiple choice questions on problem solving and gear mechanisms such as: (1) In a three-gear system (Driver $\rightarrow$ Idler $\rightarrow$ Driven), how does the driven gear's rotation direction compare to the driver gear? (2) In a robotic arm designed for precision manufacturing, why are the joints often equipped with a mechanism that allows connected segments to move at different rates?

3.4 Participants

Twenty engineering students were recruited for the study, including first and fourth year undergraduate students, as well as graduate students. This wide range of participants was chosen to study the impact of educational levels on the movement behaviours, an investigation outside the scope of this article.

Participation was voluntary with monetary compensation, and written informed consent was obtained in accordance with institutional ethics committee guidelines; an information sheet was provided to each participant. All participants' were right-handed.

3.5 Experimental Set-Up and Data Collection Protocol

Each participant completed one session, conducted in a controlled experimental setup. A session involved taking pre- and post-tests and performing the assembly task while we captured multi-modal data. As the goal is to study students' hand movements, data collected included streams of colour and depth frames from three fixed Intel RealSense D435i RGB-D cameras, positioned to capture the participant's posture, facial expressions, hand movements, and interactions with the gearbox. The cameras were mounted on a custom aluminium-extrusion frame fixed to a perforated optical breadboard: a top-down camera centered over the workspace, a mid-height oblique camera aimed at the hands, and a low front camera capturing the participant's face and posture. The work surface was a small table covered with a black mat bearing four ArUco markers at the corners to standardise the workspace coordinate frame and the placement of the parts and resting hands. Camera feeds were routed to a desktop workstation beside the worktable. Viewpoints and framing were held constant across participants, and a brief preview was performed to verify framing, focus, and exposure before each trial. The arrangement of cameras and workspace elements is summarised in Figure 1.

The three stages of the protocol were pre-test, assembly task, and post-test. 10 minutes were given for each stage. Hand placement before the start was specified, clear 'start' and 'stop' cues marked trial boundaries, and no feedback was provided until the post-test was complete. The entire protocol was completed in a single sitting.

Recordings and associated logs were de-identified and transferred immediately to secure external storage between

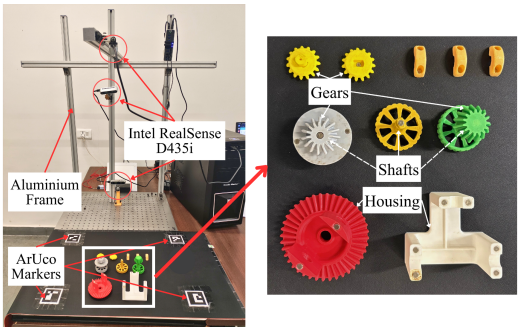

Figure 1: Experimental setup showing the camera arrangement and workspace elements.

sessions. Initially, recordings were saved as Intel RealSense .bag files to retain both RGB and depth information. To reduce storage requirements, the pipeline was later switched to saving colour and depth frames separately, with colour stored as .png and depth as .tiff. All streams were captured at 640×480 resolution and 30 fps, ensuring consistent and comparable behavioral traces while minimizing procedural variance.

3.6 Data Analysis

Data Preprocessing: Data from four participants was deemed unusable for reasons such as, file corruption and hand missing in the first frame. The remaining 16 participants' data was analysed as described here. The saved frame data obtained through the pre-processing described above formed the basis for all subsequent analysis. Regardless of whether recordings were originally captured as .bag files or directly as frame folders, the data available for analysis consisted of synchronised colour and depth frames standardised across participants. These frame datasets were then used as input for extracting hand keypoints, which served as the primary representation of participant actions during the assembly task.

Model Validation: Hand keypoint extraction was performed using the MediaPipe Hand model (Zhang et al. 2020), which predicts the positions of 21 landmarks for each hand. It is a convolutional neural network hand tracking pipeline that tracks 2D and 3D landmarks. Before applying the model across the dataset, two validation checks were performed. First, the detection percentage was calculated as the proportion of frames in which one or both hands were missing, which amounted to 4.69% of frames with both hands missing and 14.43% of frames with the right hand missing, which was deemed reasonable to proceed. Second, we validated the accuracy of MediaPipe for keypoint detection on the InterHand 2.6M (Moon et al. 2020) dataset, which has images of hands doing different kinds of actions similar to what we expect to see in our data. The MediaPipe predicted landmark coordinates of images from this dataset were compared against their ground truth annotations to assess accuracy. Combining visual inspection with RMS error analysis, we concluded that keypoints which were grossly visible across the dataset showed an offset range of 0–7 cm from the ground truth for 75% of the validation sample, which is sufficient accuracy for our purposes. Finally, we identified landmarks in our dataset that showed the most stable detection via visual inspection and found them to be landmarks 0 to 4 (spanning the tip of the thumb to wrist) on each hand. So we only used these five landmarks in the rest of the analysis. These validation steps confirmed the suitability of the MediaPipe for analysing our collected data.

Hand Keypoint Extraction: Once validated, the model was deployed on the complete dataset. For each frame, the predicted landmarks were saved in a file containing the flattened (x, y) pixel coordinates of all 21 landmarks for both hands, while annotated colour frames with hand mesh overlays were stored for visual inspection. During this stage, occasional label switching errors were observed, where the model identified both hands as left or right. To mitigate this, the handedness confidence scores provided by MediaPipe were incorporated, and only labels with a confidence above 80% were accepted as reliable, though mislabeling persisted in $<5\%$ of the frames, which were dropped from further analysis.

The 2D pixel coordinates were then combined with depth frames and camera intrinsics to convert the 2D coordinates into 3D world coordinates in mm. This conversion resulted in the 3D landmark coordinates in mm, aligned to the camera reference frame. These coordinates provided the standardised data on which all subsequent movement analysis was performed. From these world coordinate data, the movement of selected landmarks was computed across frames. For each participant, total task duration, cumulative movement of the specified landmarks, and average speed of movement, calculated as movement divided by the time spent on task, were computed.

To capture temporal variation in movement, the data were segmented using two complementary approaches. In one, the total task duration was divided into three equal segments, and movement and speed values were calculated separately for each. In the other, movement was computed for 10 s intervals, and these values were aggregated to construct cumulative movement profiles. The resulting profiles were visualised as line plots with piecewise regression fits, which provided insight into temporal changes in movement and highlighted differences between participant groups.

Although the analysis pipeline itself was identical across participants, two separate classification schemes were applied: (a) experts and novices and (b) learners and non-learners. In this paper, we will be reporting on the findings

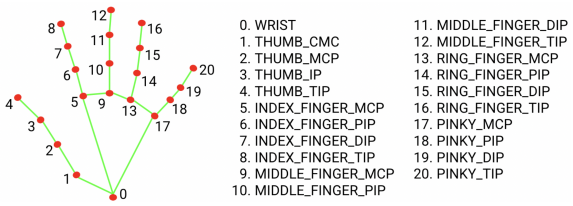

Figure 2: Hand landmarks (Google AI Edge 2019)

of the learner versus non-learner classification as the motivation for this work was to arrive at computationally tenable metrics of learning.

4 Findings

4.1 Differences in Learning

We computed learning gains as the difference between participants' scores on their pre- and post-tests. Participants with positive learning gains were classified as learners, and those with no or negative learning gain as non-learners. Six participants were classified as learners, and 10 as non-learners. We found statistically significant differences between the learning gains of the learners and the non-learners group (mean=2.67, SD=1.033; mean=-0.9, SD=0.994, respectively, p <0.001, effect size = -3.537) (see Figure 3).

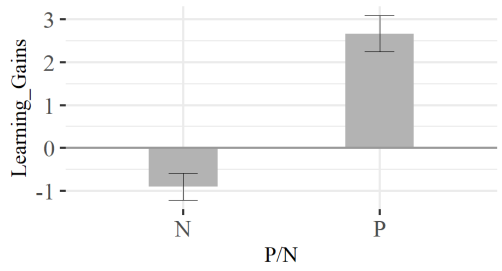

Figure 3: Learning gains of learners (P) and non-learners (N)

4.2 Differences in Movement

We saw significant differences in the total movement of hand landmarks 0 and 1 (Table 1) of right hand where learners showed significantly lower movement than non-learners with large effect sizes (1.275 and 1.239, respectively). Landmark 2 of the right hand showed a similar but marginally significant (p = 0.07) trend with a large effect size of 0.806. For landmark 3 of the right hand, and landmarks 0-3 of the left hand, mean values of the two groups show similar trends, and the differences are non-significant with small to moderate effect sizes. However, the median of the two groups (non-parametric data) show an opposite trend for landmark 4 of each hand, with learners having a higher median than non-learners, and small effect sizes.

4.3 Differences in Speed of Movement:

Considering the speed of hand movement during the task, Table 2 shows that the speed of movement for landmarks 0 and 1 of the right hand shows a trend of being marginally significantly lesser for learners than for non-learners with large effect sizes (0.903 and 0.855, respectively). A similar trend is observed in landmark 2 of both hands and landmark 1 of the left hand, however, with small to moderate effect sizes. Whereas, landmark 4 for the right hand shows a higher speed of movement among learners than non-learners with a large effect size (-1.156). A similar trend is observed for landmark 3 of both hands, as well as landmarks 0 and 4 for the left hand, but with small to moderate effect sizes.

Land-mark	non-Learners [a] Mean(SD)	Learners[b] Mean(SD)	p	Effect Size
R0	86,371 (15,771)	66,239 (15,830)	**0.014**	1.275
R1	86,948 (16,274)	67,879 (13,658)	**0.015**	1.239
R2	92,874 (13,167)	81,413 (15,927)	0.07	0.806
R3	96,518 (14,085)	91,601 (16,981)	0.27	0.324
R4	105,400[c]	118,500[c]	0.8	-0.443
L0	92,764 (23,131)	89,761 (31,133)	0.414	0.114
L1	95,372 (22,158)	87,380 (30,779)	0.277	0.312
L2	93,110 (16,822)	84,756 (26,252)	0.224	0.404
L3	95,898 (16,110)	88,278 (25,662)	0.237	0.38
L4	108,100[c]	113,100[c]	0.319	0.249

[a] N = 10. [b] N = 6. [c]: Median

Table 1: Analyses of differences in movement (mm) of right and left hands between Learners and non-Learners

4.4 Differences Across Time

Differences Across Three Equal Segments: Next, we compared movement and speed across the three time segments to explore trends across the problem-solving process. We performed repeated measures ANOVA with learners and non-learners as between subjects factors. Overall, there was no difference between learners and non-learners or any interaction effects. However, we found that the total movement of right hand landmarks 2, 3, and 4 drops significantly with time (p = 0.024; p = 0.013; p = 0.024 respectively). Post-hoc Bonferroni-adjusted comparisons indicated that only the first and third time segments for landmark 3 were statistically different. Similarly, speed of landmarks 2, 3, and 4 of right hand drops significantly with time (p = 0.014; p = 0.006; p = 0.013 respectively). Bonferroni-adjusted comparisons show statistical differences between first and third time segments for landmarks 2 and 3. Similarly, for the left hand, total movement of landmarks 1, 2, and 3 drops significantly with time (p = 0.012; p= 0.018; p = 0.02 respectively), while the speed of landmarks 0, 1, 2, and 3 drops

Land-mark	non-Learners Mean(SD)[a]	Learners Mean(SD)[b]	p	Effect Size
R0	143.5 (26.13)	121 (22.53)	0.051	0.903
R1	144.5 (27.01)	124 (16.8)	0.06	0.855
R2	154.3 (21.87)	149.2 (20.98)	0.327	0.236
R3	160.4 (23.38)	168.4 (24.82)	0.739	-0.339
R4	183.2 (25.03)	218.4 (38.44)	0.979	-1.156
L0	154.1 (38.28)	161.6 (42.24)	0.641	-0.19
L1	158.4 (36.72)	157 (41.43)	0.471	0.038
L2	154.7 (27.84)	152.5 (33.14)	0.444	0.074
L3	159.3 (26.63)	159.4 (32.45)	0.501	-0.002
L4	176.3 (25.91)	181 (38.53)	0.614	-0.153

[a] N = 10. [b] N = 6

Table 2: Analyses of differences in speed of movement (mm/s) of right and left hands between Learners and non-Learners

significantly across time segments (p = 0.047; p = 0.008; p = 0.012; p = 0.013 respectively). Post-hoc pairwise comparisons using the Bonferroni correction show that movement and speed of only first and third segments were statistically different. Thus, we observe a trend of movement and speed being lower in the third time segment compared to the first one.

Differences in Three Equal Segments: Next, we dig deeper to understand the differences in right-hand movements (refer to Table 3) and speed of movement (Table 4) of learners and non-learners in three equal time segments of the task. In the interest of space, we only report findings of the right hand landmarks where we observed significant differences between learners and non-learners; the left hand landmarks showed no significant differences across time between the two groups. In the first part of the task, movement in landmarks 0 and 1 of learners shows a tendency to be significantly lesser than non-learners with large effect sizes (0.951 and 0.938, respectively). Landmarks 2 and 3 also show similar trends with medium effect size of 0.717 and 0.403 (respectively), though non-significant. Landmarks 0 to 3 of part 2 follow the same pattern as those of part 1, first two show a tendency to be significant with large effect sizes (0.952 and 0.939, respectively), whereas the latter two landmarks have small to medium effect sizes (0.36 and 0.054, respectively). For landmarks 0 and 1 in part 3, we found significantly lesser movement in learners than non-learners with large effect sizes (1.7 and 1.586, respectively). A similar trend is noted for landmark 2 with large effect size (1.024), with a tendency to have significant differences between groups. However, in landmark 4 for parts 1 through 3, movement of learners is more than that of non-learners with small to medium effect sizes (-0.194, -0.726, and -0.305, respectively), although non-significant. With respect to speed of movement, landmarks 0 and 1 for parts 1 and 2 consistently show that speed of movement is less for learners than non-learners, though non-significant, with medium effect sizes (see Table 4). Moreover, for landmarks 0 and 1 in part 3, the speed of movement of learners is significantly lesser than that of non-learners with large effect sizes (1.323 and 1.199, respectively). Contrary to this, for landmarks 3 and 4 in parts 1 through 3, the speed of movement is more in learners than non-learners with small to large effect sizes, though non-significant for all except one landmark. Concretely, for landmark 4 in part 2, the speed of movement of learners is significantly more than non-learners with a large effect size (-1.453).

Differences in Cumulative Movement Across Time: Figures 4 and 5 show the cummulative movements of the landmarks 0 and 1 of the right hand (which showed differences between learners and non-learners in the analysis above) across the duration of the task. We split the graph with six segments, and the line fits for each of them are also indicated on the graph. The slope of each line represents the speed of movement of each landmark. We observe

Land-mark	non-Learners Mean(SD)[a]	Learners Mean(SD)[b]	p	Effect Size
1-R0	29,844 (7,205)	22,680 (8,086)	**0.043**	0.951
1-R1	30,440 (7,823)	23,558 (6,379)	**0.045**	0.938
1-R2	33,552 (6,381)	28,931 (6,569)	0.093	0.717
1-R3	34,952 (6,532)	32,344 (6,364)	0.224	0.403
1-R4	40,047 (7,029)	41,581 (9,314)	0.643	-0.194
2-R0	27,876 (5,594)	22,836 (4,704)	**0.043**	0.952
2-R1	28,554 (6,448)	23,137 (4,294)	**0.045**	0.939
2-R2	29,543 (5,777)	27,515 (5,352)	0.248	0.36
2-R3	31,171 (5,768)	30,873 (5,168)	0.459	0.054
2-R4	35,305 (5,118)	39,763 (7,633)	0.909	-0.726
3-R0	28,620 (4,591)	20,717 (4,751)	**0.003**	1.7
3-R1	27,933 (3,913)	21,178 (4,820)	**0.004**	1.586
3-R2	29,759 (3,677)	24,951 (6,118)	**0.034**	1.024
3-R3	30,370 (4,738)	28,377 (6,839)	0.25	0.357
3-R4	34,859 (6,946)	37,199 (8,860)	0.718	-0.305

[a] N = 10. [b] N = 6

Table 3: Analyses of differences in movement (mm) in 3 equal time segments of right hand between Learners and non-Learners

Land-mark	non-Learners Mean(SD)[a]	Learners Mean(SD)[b]	p	Effect Size
1-R0	149.1 (36)	125 (41.43)	0.12	0.633
1-R1	152.1 (39.13)	129.9 (31.58)	0.131	0.604
1-R2	167.6 (32.01)	159.9 (32.45)	0.326	0.239
1-R3	174.6 (32.78)	180 (37.17)	0.618	-0.157
1-R4	200.1 (35.34)	231.1 (51.28)	0.914	-0.742
2-R0	139.2 (27.963)	124.6 (12.941)	0.126	0.617
2-R1	142.6 (32.267)	126.5 (8.382)	0.128	0.613
2-R2	147.6 (28.914)	151.1 (17.706)	0.604	-0.139
2-R3	155.7 (28.872)	170 (16.845)	0.855	-0.567
2-R4	176.4 (25.69)	220.2 (36.941)	0.993	-1.453
3-R0	142 (22.57)	113.4 (19.87)	**0.011**	1.323
3-R1	138.6 (19.21)	115.7 (18.96)	**0.018**	1.199
3-R2	147.6 (17.88)	136.5 (25.87)	0.162	0.527
3-R3	150.7 (23.21)	155.3 (28.02)	0.638	-0.186
3-R4	173 (34.34)	204 (37.4)	0.944	-0.876

[a] N = 10. [b] N = 6

Table 4: Analyses of differences in speed of movement (mm/s) in 3 equal time segments of right hand between Learners and non-Learners

that learner landmarks consistently moved slower than non-learner landmarks. Further we observe that while the learners slow down towards the end, non-learners speed increases or stays the same towards the end.

5 Discussion and Conclusions

In this article, we explore the potential of computational approaches to identify hand movement related features that can distinguish between students who learned and those who did not through a making-based learning task. We performed an exploratory study with 16 engineering students who assembled a differential gearbox in a maximum of 10 minutes, and evaluated their knowledge related to gears in pre and post tests. We found that six participants had a positive learning gain, while 10 had a negative or no learning

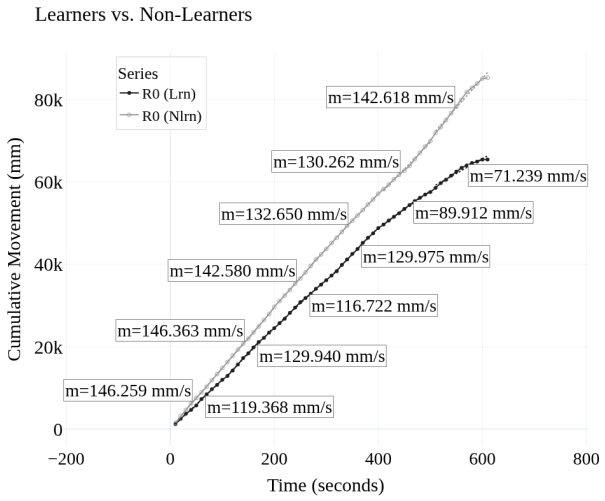

Figure 4: Cummulative movement of Landmark 0, right hand

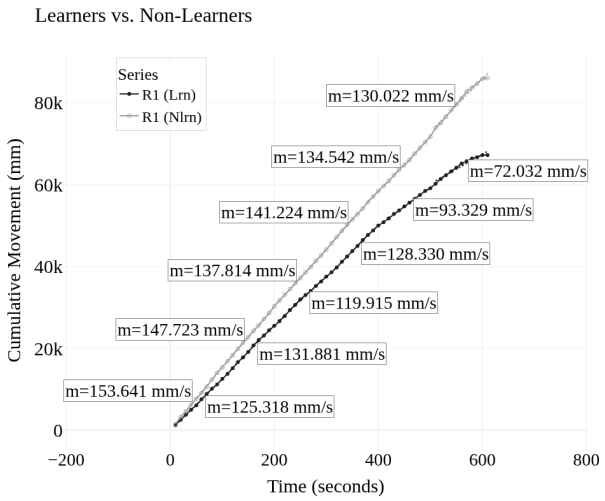

Figure 5: Cummulative movement of Landmark 1, right hand

gain. Categorizing these two groups as learners and non-learners respectively, we explored the movement of the chosen five hand landmarks, in three steps. We hypothesized based on literature of expert-novice hand movement differences that learners would have lesser and slower movement compared to non-learners. First, we examined differences between learners and non-learners on their total movement and found that right hand landmarks 0 and 1, which correspond to the wrist and base of the thumb, show significantly different movement between learners and non-learners, supporting our hypothesis. Concretely, we found that learners' landmarks moved lesser than non-learners', suggesting that learners performed lesser gross movement of their hands and thumb. This trend, though non-significant, in fact repeats across all landmarks except landmark 4, which corresponds to the tip of the thumb, that shows higher movements among learners than non-learners, and disagrees with our hypothesis. This suggests than learners show more fine movement of their thumb than novices.

Next we examined the differences between learners and non-learners on the metric of speed of movement. We found that there were nearly statistically significant differences between the two groups in their speed of movement of right hand landmarks 0 and 1, with the learners' speed of movement being lower than non-learners', supporting our hypothesis. Further we found that the speed of right hand landmark 2 also followed the same trend; however landmarks 3 and 4 which correspond to the top segment of the thumb, show the opposite trend, disagreeing with our hypothesis. Again this suggests that while learners' gross movements (wrist and base of thumb) are slower, their finer movements enabled by the top segment of the thumb are faster. These findings partly agree with previous work related to the differences between expert and novice watchmakers (Yao and Billard 2020) and surgeons (Uemura et al. 2014) where researchers found that experts movements were smoother. However, our finding that the trend of the finer movement is reversed, merits further investigation, particularly in relation to the nature of movement being performed during this task, by these hand landmarks. Finally, we found that while learners slowed down towards the end of the tasks, non-learners speed increased towards the end. This suggests a rushed effort, possibly stemming from frustration or awareness that their task time was ending.

Our findings have implications for the design of learning support for making-based learning. Specifically, our findings suggest that *low cost* metrics such as quantity and speed of movement of even a few hand landmarks, extracted from a RGB-D camera feed can indicate whether students are learning or not. This presents the possibility for automated interventions which detect learners speed and intervene when appropriate, or share this information with a teacher who can then intervene.

Our study has some limitations; first, our sample size is limited to 16 learners who did a short task (maximum of 10 minutes). Although we did obtain significant differences and large effect sizes in certain cases, the lack of significant differences in other cases suggests that our sample size is not sufficient to detect the effect. Second, our findings apply to a specific type of task, namely the gearbox assembly, which requires specific kinds of movements that may not extend across other tasks. Third, our findings are based on five landmarks which showed stable detection; owing to occlusion other landmarks could not be stably detected. It limits conclusions related to the *nature* of finer movement.

In conclusion, despite these limitations, our work presents, to the best of our knowledge, the first attempt to leverage computer vision–based identification of hand keypoints, to develop metrics for learning during making tasks, revealing distinct patterns in movement quantity and speed. These insights offer the potential for scalable, automated assessment of making-based learning and our future work includes collecting more data and expanding the types of metrics that can be used to effectively differentiate learning.

Acknowledgements

The authors acknowledge funding support from the Scheme for Promotion of Academic and Research Collaboration (SPARC) program (Project No. P3727) of the Ministry of Education, Government of India.

References

Alibali, M. W.; Spencer, R. C.; Knox, L.; and Kita, S. 2011. Spontaneous gestures influence strategy choices in problem solving. *Psychological science*, 22(9): 1138–1144.

Andolfi, V. R.; Di Nuzzo, C.; and Antonietti, A. 2017. Opening the mind through the body: The effects of posture on creative processes. *Thinking Skills and Creativity*, 24: 20–28.

Andrews, M. E.; Borrego, M.; and Boklage, A. 2021. Self-efficacy and belonging: The impact of a university makerspace. *International Journal of STEM Education*, 8(1): 24.

Carfi, A.; Motolese, C.; Bruno, B.; and Mastrogiovanni, F. 2018. Online human gesture recognition using recurrent neural networks and wearable sensors. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 188–195. IEEE.

Chandrasekharan, S.; Sinha, R.; and Date, G. 2025. Making cognition: a mechanism account of the way humans develop the ability to build and alter their environments. *Synthese*, 206.

Davidsen, J.; and Christiansen, E. 2014. Mind the Hand: A Study on Children's Embodied and Multimodal Collaborative Learning around Touchscreens. *Designs for Learning*, 7(1): 34–53.

Davidsen, J.; and Ryberg, T. 2017. "This is the size of one meter": Children's bodily-material collaboration. *International Journal of Computer-Supported Collaborative Learning*, 12(1): 65–90.

Davis, R. L.; Schneider, B.; Rosenbaum, L. F.; and Blikstein, P. 2024. Hands-on tasks make learning visible: A learning analytics lens on the development of mechanistic problem-solving expertise in makerspaces. *Educational technology research and development*, 72(1): 109–132.

Echeverría, V.; Avendaño, A.; Chiluiza, K.; Vásquez, A.; and Ochoa, X. 2014. Presentation skills estimation based on video and kinect data analysis. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, 53–60.

Google AI Edge. 2019. MediaPipe Hand Landmarker. https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker. Accessed: 2025-08-19.

Lyu, Q.; Chen, W.; Liu, S.; John Gerard Heng, K. H.; Su, J.; and Wang, Y. 2025. Hands-on Consensus Building: Leveraging Deep Learning Models to Unveil Hand Gestures in Consensus-Building Discourses. *Cognition and Instruction*, 43(1-2): 33–68.

Lyu, Q.; Chen, W.; Su, J.; Heng, K. H. J. G.; and Liu, S. 2023. How peers communicate without words-an exploratory study of hand movements in collaborative learning using computer-vision-based body recognition techniques. In *International Conference on Artificial Intelligence in Education*, 316–326. Springer.

Moon, G.; Yu, S.-I.; Wen, H.; Shiratori, T.; and Lee, K. M. 2020. InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image. In *European Conference on Computer Vision (ECCV)*.

Nasir, J.; Kothiyal, A.; Bruno, B.; and Dillenbourg, P. 2021. Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. *International Journal of Computer-Supported Collaborative Learning*, 16(4): 485–523.

Niu, Z.; Lu, K.; Xue, J.; Qin, X.; Wang, J.; and Shao, L. 2024. From methods to applications: A review of deep 3d human motion capture. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(11): 11340–11359.

Noël, R.; Miranda, D.; Cechinel, C.; Riquelme, F.; Primo, T. T.; and Munoz, R. 2022. Visualizing collaboration in teamwork: A multimodal learning analytics platform for non-verbal communication. *Applied Sciences*, 12(15): 7499.

Norton, R. L. 2009. *Kinematics and dynamics of machinery*. McGraw-Hill New York.

Oudah, M.; Al-Naji, A.; and Chahl, J. 2020. Hand gesture recognition based on computer vision: a review of techniques. *journal of Imaging*, 6(8): 73.

Pavlakos, G.; Shan, D.; Radosavovic, I.; Kanazawa, A.; Fouhey, D.; and Malik, J. 2024. Reconstructing hands in 3d with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9826–9836.

Riquelme, F.; Munoz, R.; Mac Lean, R.; Villarroel, R.; Barcelos, T. S.; and de Albuquerque, V. H. C. 2019. Using multimodal learning analytics to study collaboration on discussion groups: A social network approach. *Universal Access in the Information Society*, 18(3): 633–643.

Schad, M.; and Jones, W. M. 2020. The maker movement and education: A systematic review of the literature. *Journal of Research on Technology in Education*, 52(1): 65–78.

Sinha, R.; and Chandrasekharan, S. 2021. Embodied learning in makerspaces. In *International Conference on Computers in Education*.

Soomro, S. A.; Casakin, H.; Nanjappan, V.; and Georgiev, G. V. 2023. Makerspaces fostering creativity: A systematic literature review. *Journal of Science Education and Technology*, 32(4): 530–548.

Sung, H.; and Nathan, M. J. 2024. Your body tells how you engage in collaboration: Machine-detected body movements as indicators of engagement in collaborative math knowledge building. *British Journal of Educational Technology*, 55(5): 1950–1973.

Turakhia, D.; Ludgin, D.; Mueller, S.; and Desportes, K. 2024. Understanding the educators' practices in makerspaces for the design of education tools. *Educational technology research and development*, 72(1): 329–358.

Uemura, M.; Tomikawa, M.; Kumashiro, R.; Miao, T.; Souzaki, R.; Ieiri, S.; Ohuchida, K.; Lefor, A. T.; and Hashizume, M. 2014. Analysis of hand motion differentiates

expert and novice surgeons. *Journal of Surgical Research*, 188: 8–13.

Winters, S.; Farnsworth, K.; Berry, D.; Ellard, S.; Glazewski, K.; and Brush, T. 2023. Supporting middle school students in a problem-based makerspace: Investigating distributed scaffolding. *Interactive Learning Environments*, 31(6): 3396–3408.

Worsley, M.; and Blikstein, P. 2018. A multimodal analysis of making. *International Journal of Artificial Intelligence in Education*, 28(3): 385–419.

Yao, K.; and Billard, A. 2020. An inverse optimization approach to understand human acquisition of kinematic coordination in bimanual fine manipulation tasks. *Biological Cybernetics*, 114: 63–82.

Zhang, F.; Bazarevsky, V.; Vakunov, A.; Tkachenka, A.; Sung, G.; Chang, C.-L.; and Grundmann, M. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*.