scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A Multisensor Dataset of South Asian Post-Graduate Students Working on Mental Rotation Tasks

Ashwin T. S. ✉, **Suraj Ranganath**, **Kabyashree Khanikar**, **Karishma Khan**, **Ramkumar Rajendran** ✉ & **Ritayan Mitra** ✉

Spatial thinking, in general, and mental rotation, in particular, have seen sustained research attention due to such abilities playing a critical role in STEM (science, technology, engineering and mathematics) learning. The recent development of sensor-based approaches to identify, understand, and measure cognition and affect opens up new possibilities to study such topics. We collected galvanic skin response, electroencephalography, screen recording, facial expressions, manual emotion logging, task performance logs (including response times, correctness, and question difficulty), gaze, and self-reports of 38 participants as they solved mental rotation tasks under various conditions, namely, (i) with no time restriction and no feedback, (ii) with no time restriction and with feedback and (iii) with time restriction and with no feedback, respectively. The availability of such a dataset will help researchers in the spatial thinking community to study interesting questions related to strategy selection, flexibility, affective response, and group differences in mental rotation tasks. Furthermore, the learning analytics community could gain valuable insights into how providing feedback might change learning and engagement during such tasks.

Background & Summary

Decades of research on mental rotation have yielded some robust yet intriguing results. Previous literature has reported higher levels of performance for male participants than for female participants in mental rotation tasks. However, there is variability in these findings, with some studies indicating higher performance in males (e.g.[1–3]), while others, such as Moen *et al.*[4], find no significant gender differences in mental rotation ability after training. Moreover, meta-analyses suggest that gender differences can be influenced by the time limits imposed during testing; the performance gap tends to diminish or disappear when strict time limits are removed[5,6]. The underlying cause of such group and/or individual differences could be related to biological[7,8], spatial anxiety[9], strategic[10], experiential[11,12], and affective[13]. More research is still warranted in this direction, notwithstanding novel suggestions, such as the co-construction of the ability itself with that of gender identity[14].

Another interesting result is observed in the variation of the strategy adopted by individuals. Holistic rotation, through which entire objects are rotated in the working memory, is considered to be less common and potentially more effective than piecemeal rotation, in which parts of the objects are rotated mentally[15–18]. Such pure rotation strategies are often used together with analytic strategies, such as global shape, perspective taking, local turns, and cube counting[3,19]. The selection (choice of a strategy) and flexibility (changing strategy to match item difficulty) have been suggested to be related to overall performance in such tests. However, selection and flexibility are still poorly understood in part due to the difficulty of obtaining strategy information from participants. As most spatial thinking studies are conducted with surveys, think-aloud, either concurrent or retrospective, and interviews, there is always a chance of imperfect recall or lack of awareness of the strategies being used by participants.

The spatial thinking research community has traditionally relied on accuracy, response time, questionnaires, think-aloud, and interview data. The role of multimodal data in understanding the complexities of cognition and affect within any context, as can be expected in mental rotation tasks, cannot be overlooked in an age where such data collection has become more accessible and efficient. Indeed, such data are increasingly being used for spatial thinking and similar contexts. For example,eye-tracking technology offers advantages over traditional

Indian Institute of Technology Bombay, IDP in Educational Technology, Mumbai, 400076, India. ✉e-mail: ashwindixit9@gmail.com; ramkumar.rajendran@iitb.ac.in; rmitra@iitb.ac.in

data collection methods, such as verbal self-reports and checklists, as the gaze data can be used to understand strategy use[18,20]. Likewise, electroencephalogram (EEG) spectral powers could provide useful insights on the approach/avoidance to such problems or the mental workload experienced during problem-solving[21]. Additionally, galvanic skin response (GSR) and facial expressions could give us insights into the affective states of participants as they answer such questions[22,23]. If such modalities are combined with traditional surveys and interviews, then we could answer some of the persistent questions that are raised by the community.

This database was designed considering the demands of the spatial thinking community where data are largely of one type and from a limited geographical location (excluding India). We have collected multimodal data from 38 participants solving mental rotation problems. The data includes eye tracking, facial features, GSR, EEG, posture and other relevant factors as expressed through a battery of survey questions and interviews. We hope this dataset will allow researchers to fully understand students' cognitive processes and affective states when performing complex spatial thinking tasks. Furthermore, we have introduced a category of questions, namely, answers with feedback, that would also inform how students strategize and perform in such tests when provided feedback (correct or incorrect) on their answers. Furthermore, the open sharing of data for spatial thinking research is expected to encourage studies looking at the generalizability and replicability of results and trigger across-culture analysis of such data.

In summary, this study aims to fill the gaps in the existing literature by developing a multimodal database for understanding the spatial thinking process in college students. By incorporating multiple modalities and considering cultural and regional differences, we can advance our understanding of spatial thinking abilities and their implications for learning and cognition. Furthermore, this dataset represents the first publicly available multimodal database in this field, offering a unique resource that combines cognitive, affective, and physiological data. This availability could play a crucial role in the development and refinement of computational models that simulate cognitive and affective processes in spatial thinking tasks, thereby advancing both theoretical and applied research.

Methods

Ethical Statements. Informed consent was obtained from all participants prior to their participation in the study. Participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits, and they voluntarily agreed to participate. Notably, explicit consent was secured from all participants for the open publication of the data, which has consequently been made available on Figshare under a CC0 license, in adherence to the obtained consent. In accordance with ethical guidelines, the data will be anonymized and shared with the public as a database to facilitate further exploration in the field. The study protocol, including the consent process and data handling procedures, has been reviewed and approved by the Institutional Review Board (IRB) committee of the Indian Institute of Technology (IIT) Bombay. The IRB approval for this study is granted under IRB number IITB-IRB/2022/006, ensuring compliance with ethical standards and the protection of participants' rights.

Study Design. The mental rotation task (MRT), developed by Shepard and Metzler[15], serves as the primary task in this study, where participants determine whether two figures are the same. To facilitate the MRT study, the open-source package PsychoPy (https://www.psychopy.org/) is employed for designing the user interface. Additionally, iMotions (https://imotions.com/) software is utilized to collect data related to eye tracking, galvanic skin response, and affect. A human observation tool[24], Data Logging and Organizational Tool (DLOT) (https://danishsshaikh.github.io/DLOT/) is employed for manual emotion logging, while a separate tablet is used for this purpose. To collect electrocochleography data, Muse hardware is utilized in conjunction with Mind monitor software (https://mind-monitor.com/). Several self-report measures are incorporated, including demographic information, spatial anxiety[25], personality traits[26], metacognitive self-regulation[27], and the NASA Task Load Index (NASA TLX)[28] to measure subjective mental workload.

After obtaining participant consent, a presurvey questionnaire and working memory test (Backward Corsi Block Tapping Test and N-back Test) are administered, with an approximate completion time of 25 minutes. Participants are then seated in front of a system with iMotions software installed. In this study, two rooms are utilized: one serves as an observation room, where a trained observer equipped with a tablet manually logs emotions using DLOT. This observation room is equipped with a one-way mirror, ensuring that participants do not have a visual reminder of the observer's presence. However, participants are informed about the observer and the purpose of the observation room.

Participants are comfortably seated in the second room, which houses a laptop with iMotions software installed. The study begins with introducing the various sensors and requesting participants to assume a comfortable position for sensor calibration. Several sensors are employed and calibrated, including a Tobii eye tracker positioned at the laptop screen's bottom. The distance from the participant's eyes to the laptop screen was maintained approximately at 60 cm by instructing participants to sit at this distance, as illustrated in Fig. 1. The study commences only when eye tracking calibration achieves a 'good' or 'excellent' status.

Additionally, a GSR (Galvanic Skin Response) sensor is attached to the participant's non-dominant hand (left hand for right-handed users, and vice versa). EEG electrodes are placed on the participant's head and calibrated for all electrodes. With all sensors properly calibrated, the study officially begins. The study's overall setup and configuration are visually represented in Fig. 1, providing an overview of the study environment.

The study consists of three categories: the first without feedback and no time restriction (Category 1), the second with feedback and no time restriction (Category 2), and the third without feedback but with time restrictions (Category 3). There are no breaks between categories, and instructions for each category are provided beforehand. Before the start of each category, participants are shown two sample questions for familiarization. In categories 1 and 2, participants have 60 seconds for 2 sample questions (Screenshot of the sample question before the start of the category 1 is shown in Fig. 2). In category 3, for 2 sample questions, participants have only 12 seconds.

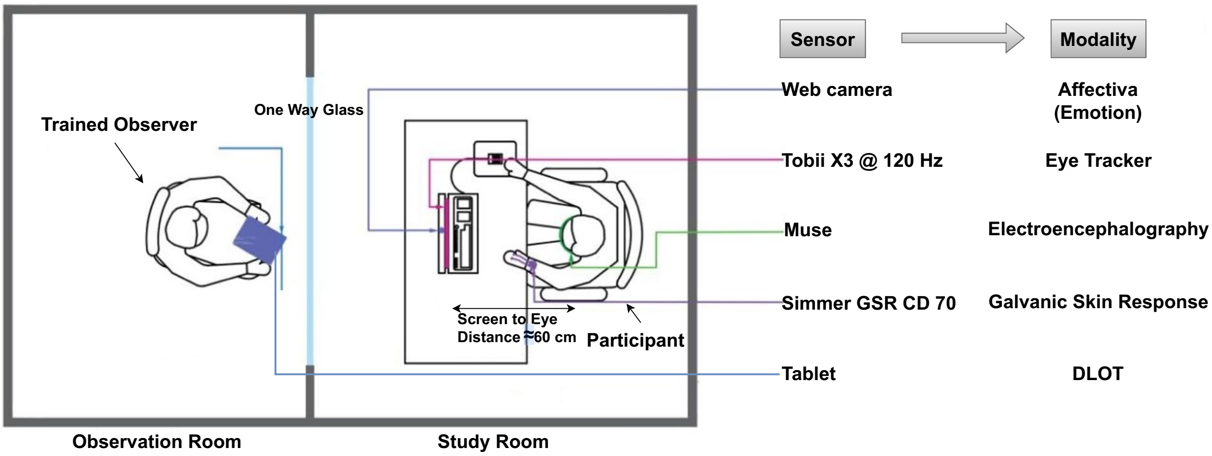

Fig. 1 Overall setup of the study.

The questions in these categories were designed to be of easy, medium, and hard difficulty levels. To achieve this, we systematically manipulated angular disparities between objects (ranging from 0 to 180 degrees). However, angular disparity alone is not the only source of difficulty in such tasks; four additional sources of difficulty were identified from psychometric literature[29,30]. These factors included configuration type (homogeneous vs. heterogeneous), peripheral arms position (parallel vs. perpendicular), occlusion (free vs. occluded), and axis of rotation (x-axis, z-axis, or multiple axes). Configuration type refers to the structural arrangement of the objects, where homogeneous configurations consist of peripheral segments with the same number of blocks (e.g., three blocks on each side), making them more challenging to differentiate and rotate mentally. In contrast, heterogeneous configurations have different numbers of blocks on the segments (e.g., two blocks on one side and three on the other), providing more visual cues for easier mental rotation[29]. Homogeneous configurations, occluded objects, and parallel arms were significantly more challenging than their heterogeneous, non-occluded, and perpendicular counterparts[29]. Additionally, rotations along the z-axis were found to be more difficult than those along the x-axis[30]. For each category, there was a mix of easy, medium, and difficult items. In Categories 1 and 2, the questions were presented sequentially based on difficulty, while in Category 3, the items were randomized. The log data provided in the shared dataset documents these details, ensuring full transparency and replicability. A sample image screenshot of the design of test items with increasing difficulty using these five factors is provided in Appendix 7 of the supplementary information. However, the full set of images from Peters & Battista (https://www.uoguelph.ca/psychology/lab/mrt)[30,31] contains a large number of items, and we used only a specific subset in this study. This subset is not publicly available and can only be accessed with permission from Michael Peters. All images used in this study have been compiled into a single document with structured metadata, including category number, image name, image, difficulty level, and match/no-match classification. Dr. Peters will provide access to this subset upon request. To obtain the exact images used in this study, researchers must reference this paper by providing the first author's name and the article title when contacting Dr. Peters at mpeters@uoguelph.ca for access.

To ensure the appropriateness of the study design, we conducted a pilot study with a sample of six participants (three males and three females). The pilot study aimed to evaluate the number of questions, time allocated per condition, and participants' responses to the tasks. The results indicated that the average response time varied between 2.4 and 3.2 seconds for questions with angular disparities ranging from 0 to 180 degrees, which is consistent with the literature[32]. The most extended completion time for any individual question was 20 seconds. Participants reported experiencing no time pressure in Categories 1 and 2, confirming that the selected number of questions and time allocation were appropriate for capturing the necessary data without causing undue fatigue. Therefore, in Categories 1 and 2, where time is not a key constraint, the extended time limit is referred to as 'no time restriction' throughout the paper.

The user interface (UI) displays two images, and participants must select whether they match or not. In the second category, a feedback component indicating the correctness of the selected option is displayed at the bottom of the UI. Although there is no time restriction for answering questions, the overall duration for categories 1 and 2 is set to 300 seconds each, with 10 questions provided. Category 3 includes 20 questions, with a duration of 120 seconds. Participants can attempt as many questions as possible from the set of 20; however, it's notable that no participant completed all 20 questions within the time limit.

The number of questions per condition was carefully chosen to align with the specific objectives of the experimental setup. For conditions 1 and 2, 10 questions were included to allow participants sufficient time to engage with the tasks without experiencing time pressure, as these categories were designed to examine performance under unconstrained conditions. In contrast, condition 3 featured 20 questions to introduce sustained time pressure, creating a scenario where participants needed to manage cognitive demands under constrained time. This differentiation allowed us to observe variations in cognitive strategies and task performance, consistent with prior research emphasizing the role of time pressure in revealing performance variations[6,30].

The justification for these choices was further supported by previous studies and validated through a pilot study. Research has demonstrated that using 10-20 questions is sufficient to conduct meaningful analyses while

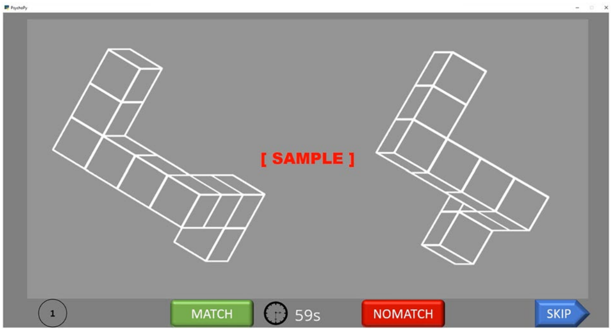

Fig. 2 Screenshot of the sample question before the start of the category 1.

minimizing participant fatigue[33]. In our pilot study, participants reported experiencing no time pressure in conditions 1 and 2, while condition 3 successfully induced the intended cognitive load. These results confirmed that the selected number of questions and time allocations effectively balanced data quality and participant engagement, ensuring the robustness of the study design.

After the completion of each category, NASA-TLX questions appear on the screen, and participants are required to complete the form before proceeding to the next category.

The study involves a total of 38 participants (22 Males and 16 Females). On average, each participant spends 25 minutes on the presurvey questionnaires, 10 minutes on the working memory test, and 25 minutes on all three categories. The multimodal data is anonymized and aligned based on the timestamp, resulting in a database size of approximately 8 GB. Figure 3 provides an overview of the database, and the subsequent subsections offer comprehensive details of the study.

Self Reports. The self-reports used in the study have various aspects, including demographics, educational background, spatial thinking abilities, anxiety levels in spatial tasks, personality traits, and learning strategies. Appendix Section 1 (Supplementary Information) presents the set of questions employed in the study. Here is a brief description of the self-reports:

Demographic Questions. Demographic information is collected from the participants through a series of questions. These questions aimed to gather data on participants' age, gender, highest qualification, study discipline, current pursuits, study discipline at present, and year of study. The set of demographic questions used in the study is given in Appendix Section 1.1.

Spatial Anxiety Questionnaire. Spatial Anxiety Questionnaire with subscales to measure three aspects of spatial anxiety were collected. The questionnaire consists of 24 items that assess various scenarios related to mental manipulation, navigation, and imagery[25]. Participants were asked to rate their level of anxiety for each item on a scale ranging from "not at all" to "very much." The questionnaire is designed to identify situations and experiences that may induce tension, apprehension, or anxiety. The final Spatial Anxiety Scale is divided into three subscales: Mental Manipulation (M), Navigation (N), and Imagery (I). Detailed instructions were provided to participants to ensure accurate responses. The questionnaires, along with the scoring guidelines, are included in Appendix Section 1.2 for reference.

BFI-10 personality trait questionnaires. BFI-10 personality trait questionnaires were used in this study, derived from the Big Five Inventory (BFI)[26]. The BFI-10 is a condensed version of the BFI, designed to accommodate situations with limited participant time. It consists of 10 items, with two representing each of the five personality dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. The BFI-10 has demonstrated favorable psychometric properties, including reliability and validity, making it a practical tool for efficiently assessing personality traits within research settings with time constraints. Scoring for the BFI-10 involves assigning reverse scores to specific items: items 1 and 5 for Extraversion, item 2 and item 7 for Agreeableness, items 3 and 8 for Conscientiousness, items 4 and 9 for Neuroticism, and items 5 and 10 for Openness to Experience. The questionnaires and the scoring scale are provided in the Appendix Section 1.3 for reference.

Metacognitive self-regulation. In our study on cognitive and metacognitive strategies, we specifically focused on the metacognitive self-regulation aspect using the Metacognitive Self-Regulation scale (Motivated Strategies for Learning Questionnaire (MSLQ))[27]. Metacognition refers to the awareness, knowledge, and control of cognition. Our emphasis was on the control and self-regulation processes rather than the knowledge aspect of metacognition. Three main processes were explored within metacognitive self-regulatory activities: planning, monitoring, and regulating. Planning activities involved setting goals and analyzing tasks to activate relevant prior knowledge, facilitating the organization and comprehension of the material. Monitoring activities encompassed tracking attention, self-testing, and questioning, aiding learners in understanding and integrating the material with their existing knowledge. Lastly, regulating involved fine-tuning and continuous adjustment of cognitive activities, allowing learners to check and correct their behavior as they progressed through a task.

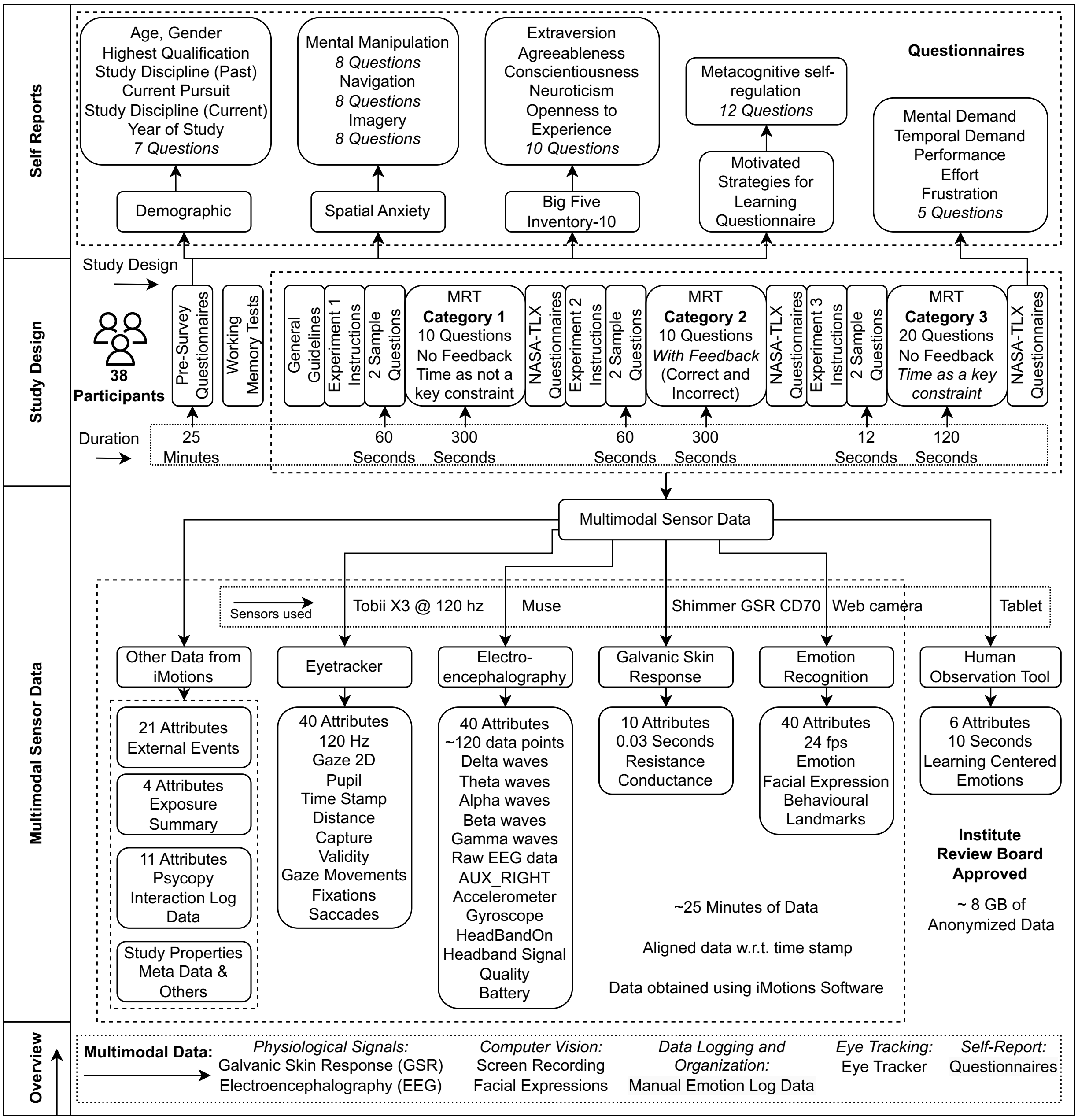

Fig. 3 Overview of the study design of the created multimodal spatial thinking study.

By examining these metacognitive self-regulation processes, we aimed to understand their impact on performance and learning outcomes in our study. The questionnaires used to measure metacognitive self-regulation are included in Appendix Section 1.4.

NASA-TLX Questionnaires. Our study used the NASA-TLX (National Aeronautics and Space Administration Task Load Index) questionnaires to assess various aspects of task load and user experience[28]. The NASA-TLX is a widely recognized and extensively validated tool for evaluating workload and subjective performance-related experiences. It comprises six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Each dimension is assessed using a 7-point Likert scale, where participants rate the intensity or impact of the specific factor on a given task. The detailed questionnaire items can be found in Appendix Section 1.5 of the study, allowing for a comprehensive understanding of the participant's subjective experiences and task load during the experimental tasks. Notably, the Physical Demand dimension was not included in our study, as the experimental tasks did not require any physical demand from the participants.

Working Memory Tests. As part of the pre-survey, two working memory tests[34,35] were administered to the participants. These tests are widely used in cognitive psychology research and provide valuable insights into different aspects of working memory. The two tests used in this study were the Backward Corsi Block Tapping Test and the N-back Test.

- Backward Corsi Block Tapping Test: The Backward Corsi Block Tapping Test (https://www.psytoolkit.org/experiment-library/backward_corsi.html) is a spatial working memory task that assesses the participant's ability to remember and reproduce a sequence of spatial locations in reverse order. In this test, participants are presented with a series of blocks arranged in a specific pattern. They are then required to tap the blocks in the reverse order of presentation. The test measures the participants' spatial working memory capacity and their ability to manipulate and hold spatial information in mind.
- N-back Test: The N-back Test (https://www.psytoolkit.org/experiment-library/nback2.html) is a widely used measure of overall working memory capacity and cognitive control. In this test, participants are presented with a sequence of stimuli, typically letters or numbers, and must indicate whether the current stimulus matches the one presented N items back in the sequence. For example, in a 2-back test, participants must indicate whether the current stimulus matches the one presented two items back. The N-back test measures the participant's ability to maintain and update information in working memory, as well as their attentional control and cognitive flexibility.

Both of these tests were administered using an online tool called PsyToolkit (https://www.psytoolkit.org/), which provides a platform for administering and scoring a wide range of psychological tests. Using PsyToolkit ensures standardized administration and scoring procedures, minimizing potential biases and increasing the reliability of the test results.

Multimodal Sensor data. *Eyetracker.* The eye tracking data in this study was collected using the Tobii X3 @ 120 Hz (https://www.tobii.com/products/discontinued/tobii-pro-x3-120) sensor integrated within the iMotions software. The data was then analyzed using the I-VT (Interval Velocity Threshold) filter, a commonly used gaze analysis method. The I-VT filter allows for identifying fixations and saccades based on velocity thresholds. The entire eyetracker data is stored in 2 csv files: The file named "EYE" contains 15 attributes related to gaze and distance. The "IVT" file contains the remaining data and has 25 attributes. The detailed summary of eyetracker data is as follows:

- Gaze 2D: Gaze 2D focuses on the coordinates of the gaze point in two-dimensional space. It includes the X-coordinate and Y-coordinate of the gaze point, which are relative to the top-left corner of the screen. The values are calculated as the average of the corresponding coordinates of the left and right eye. Additionally, it includes measurements of pupil diameter for both eyes, further enriching the gaze data.
- Distance: The estimated distance between the eye-tracker and the eyes is given by distance. The distance values are interpolated when missing and are represented as the average distance from the eye-tracker to the left and right eye.
- Gaze Movements: Gaze Movements focuses on analyzing the movement of the gaze. It includes two measurements: gaze velocity and gaze acceleration. Gaze velocity represents the angular velocity of the eyes at the current sample point, indicating how fast the eyes are moving. Gaze acceleration represents the angular acceleration of the eyes at the current sample point, indicating the rate of change in the eyes' speed.
- Fixations: Fixations are periods when the gaze remains relatively stable. It includes several measurements related to fixations, such as the fixation index (counting from the start of the recording or stimulus), X and Y coordinates of the fixation centroid (the average position of gaze points during the fixation), start time, end time, duration of the fixation, and dispersion (how spread out the gaze points within the fixation are).
- Saccades: Saccades are rapid eye movements between fixations. It includes measurements such as the saccade index (counting from the start of the recording or stimulus), start time, end time, duration of the saccade, saccade amplitude (angular distance traveled by the eyes), peak velocity (maximum speed during the saccade), peak acceleration (maximum increase in speed), peak deceleration (maximum decrease in speed), and direction of the saccade indicated as clockwise angles. The complete details of the eye tracking data are given in Appendix Section 2.

Electroencephalography. The Muse EEG headband (https://choosemuse.com/products/muse-2-x-poosh) was used to collect EEG data, and the obtained data is stored in a CSV file. The Muse EEG headband has four dry electrode channels located at the positions AF7, AF8, TP9, and TP10. Specifically, AF7 and AF8 are located on the forehead, while TP9 and TP10 are positioned on the ears. This setup corresponds to the international 10-20 system for electrode placement[36]. The data consists of various attributes that represent different types of recorded information. Here is a detailed explanation of each component:

- TimeStamp: The time at which the data point was recorded.
- Delta waves: Power measurements of delta waves at specific electrode locations on the scalp (Delta_TP9, Delta_AF7, Delta_AF8, Delta_TP10).
- Theta waves: Power measurements of theta waves at specific electrode locations (Theta_TP9, Theta_AF7, Theta_AF8, Theta_TP10).
- Alpha waves: Power measurements of alpha waves at specific electrode locations (Alpha_TP9, Alpha_AF7, Alpha_AF8, Alpha_TP10).
- Beta waves: Power measurements of beta waves at specific electrode locations (Beta_TP9, Beta_AF7, Beta_AF8, Beta_TP10).

- Gamma waves: Power measurements of gamma waves at specific electrode locations (Gamma_TP9, Gamma_AF7, Gamma_AF8, Gamma_TP10).
- Raw EEG data: Raw EEG readings from specific electrode locations (RAW_TP9, RAW_AF7, RAW_AF8, RAW_TP10).
- AUX_RIGHT: Auxiliary data recorded from a sensor placed on the right side of the headband.
- Accelerometer: Accelerometer readings measuring head movement along the X, Y, and Z axes (Accelerometer_X, Accelerometer_Y, Accelerometer_Z).
- Gyroscope: Gyroscope readings measuring rotational movement along the X, Y, and Z axes (Gyro_X, Gyro_Y, Gyro_Z).
- HeadBandOn: Indicates whether the headband was worn or not during the recording.
- Headband Signal Quality: Signal quality measurements at specific electrode locations (HSI_TP9, HSI_AF7, HSI_AF8, HSI_TP10).
- Battery: Battery level of the Muse headband.

The database includes a CSV file labeled 'EEG,' encompassing details across all 40 attributes. These attributes provide valuable insights into brainwave activity, raw EEG data, sensor readings, and headband status, allowing for analysis of mental states, head movement, and other relevant information during the EEG session. The visualization of the EEG data using the Mind Monitor software is provided in Appendix Section 3.

Galvanic Skin Response. The study uses measurements and parameters associated with the Shimmer GSR CD70 device (https://shimmersensing.com/product/shimmer3-gsr-unit/). The detailed summary of GSR data is as follows:

Galvanic Skin Response. GSR measurements reflect changes in skin conductance and resistance. The two parameters discussed are GSR Resistance, representing the skin resistance, and GSR Conductance, representing the skin conductance. These measurements provide insights into the autonomic nervous system's activity and emotional arousal levels.

Other. This section includes additional parameters and timestamps associated with the Shimmer GSR CD70 device. The parameters are as follows:

- SampleNumber: A numerical value representing the sample number of the recorded data.
- Timestamp RAW: The timestamp (unitless) provided by the Shimmer device when the data was sampled.
- Timestamp CAL: The calibrated (CAL) timestamp provided by the Shimmer device when the data was sampled.
- System Timestamp: The timestamp provided by the Shimmer SDK system when the sample arrived at the computer, allowing synchronization with other data sources or events.
- VSenseBatt RAW: The raw battery voltage data (unitless) of the Shimmer device.
- VSenseBatt CAL: The calibrated battery voltage data representing the actual battery voltage level.
- GSR RAW: The raw galvanic skin response data (unitless), which measures the skin's electrical conductance.
- Packet reception rate: The percentage of received data packets (in raw form), indicating the reliability of data transmission. A value lower than 100 indicates that some packets were dropped or not received.

These parameters and measurements provided by the Shimmer GSR CD70 device allow for the monitoring and analysis of galvanic skin response. A detailed description of these 10 attributes is provided in Appendix Section 4. The CSV file named 'GSR' in the database comprises all these 10 attributes.

Emotion Recognition. The facial features data is collected by iMotions using the Affectiva AFFDEX software development kit (SDK) for emotion recognition. The complete emotion data is stored in the database's CSV file named 'TIVA.' Here is a detailed summary of each section:

Affective AFFDEX. The Affective AFFDEX file contains 40 attributes grouped into four categories: Emotion, Facial Expression, Behavioral, and Other. These attributes provide information about various aspects of facial expressions and emotional states. Here are the details of each category:

- Emotion: This category includes attributes related to specific emotions expressed by the face, such as Anger, Contempt, Disgust, Fear, Joy, Sadness, Surprise, Engagement, Valence, Sentimentality, Confusion, and Neutral.
- Facial Expression: These attributes describe different facial movements and expressions, including Attention, Brow Furrow, Brow Raise, Cheek Raise, Chin Raise, Dimpler, Eye Closure, Eye Widen, Inner Brow Raise, Jaw Drop, Lip Corner Depressor, Lip Press, Lip Pucker, Lip Stretch, Lip Suck, Lid Tighten, Mouth Open, Nose Wrinkle, Smile, Smirk, Upper Lip Raise, Blink, and BlinkRate.
- Behavioral: This category includes attributes that capture behavioral aspects, such as Pitch (up/down rotation of the face), Yaw (left/right rotation of the face), Roll (tilt-rotation of the face), and Interocular Distance (distance between the eyes).
- Other: This component, named SampleNumber, provides a unique identifier for each sample.

The Affectiva AFFDEX SDK, used in conjunction with iMotions, applies computer vision algorithms and machine learning techniques to analyze real-time facial expressions. The collected data provides detailed insights into emotions, facial expressions, behavioral cues, and specific facial landmarks. The complete details of all the 40 features of AFFDEX SDK is provided in Section 5 of the Appendix.

Logdata from Psychopy (Task Performance Log). This study used PsychoPy, an open-source package commonly used for experiment UI design in psychology and neuroscience research, to create the experiment UI. The experimental data was saved in a CSV file labeled 'PSY' within the database, encompassing several attributes pivotal for analysis and interpretation.

The CSV file contained the following attributes:

- Key: A unique key generated using the combination of category and question numbers, including the sample questions.
- Category: Indicates which category the data belongs to, with values 1, 2, or 3.
- QuestionNumber: Represents the question numbers, including the sample questions used in the categories.
- match Ornomatch: Denotes whether the participant's response matched or did not match the given stimulus.
- Difficulty: Indicates the difficulty level of each question, categorized as Easy, Medium, or Difficult.
- Verdict: Specifies whether the participant's response was correct or incorrect.
- ResponseTime: The time taken by the participant to respond to each question.
- routineStart: The timestamp at the beginning of the routine.
- routineStamp: The timestamp indicates the time when the routine occurred.
- routineEnd: The timestamp at the end of the routine.
- Cat2FeedbackTime: Represents the feedback time for category 2 in the experiment.

Data Logging. The learning-centred emotions in a classroom setting are manually observed using an open-source application called Data Logging and Organizational Tool (DLOT)[24] at 10-second intervals. The DLOT application allows human observers to log data and provides a timestamp for each observation. It is a flexible tool that can be customized to work with different operating systems and offers various options for saving and sharing data. The tool is open-source and available on GitHub (https://github.com/danishsshaikh/DLOT/).

In the DLOT application, the learning-centred emotions[37–39] considered are as follows: engaged, neutral, boredom, frustrated, confused, and delighted. The human observer's task is to observe the participant and select the appropriate affective state from the available options every 10 seconds. The application includes a timer that displays the time remaining, ensuring that the observer selects an affective state within the designated time frame.

The logged data is stored in .xlsx format, which includes information such as the timestamp of each observation, participant ID, and the selected affective state. The application automatically records the timestamp and user details when selecting an affective state. The database includes the details stored in the CSV file named 'DLOT'.

In the study, two different observers were involved, with each observer assigned to observe a specific participant. Before the study commenced, the observers underwent training, which included a standard training procedure. During the training, the observers were familiarized with the standard definitions of learning-centred affective states and were trained to ensure consistency in their observations.

An inter-observer reliability test was conducted to assess the observers' reliability. In the intra-observer reliability test, the observers were individually instructed to observe videos of students obtained from the DAiSEE dataset[40]. They compared their selections of affective states with the existing annotations in the dataset and discussed any disagreements or discrepancies they encountered. The observers worked collaboratively to reach a consensus.

Following the initial training and reliability assessment, a pilot study was conducted with four participants from the spatial thinking study. These participants were different from the 38 participants in the created database. During the pilot study, both observers independently observed the same participants and selected the appropriate affective state using the DLOT application at 10-second intervals. The intra-rater reliability, measured using Cohen's kappa coefficient, was determined to be 0.91, indicating a high level of agreement between the observers. Throughout the study, each participant was observed by only one observer as they were already trained, and their interrater reliability is very high.

Data Records

The entire multimodal database, accessible for public use, is shared through fig**share**[41], with the dataset structured as follows: an Excel file titled "Pre-survey for MMA Research" and exposure summary of all 38 participants can be found in the main repository. In the "STData" folder, there are 38 subfolders, each corresponding to one of the 38 participants in the study. Within these participant-specific subfolders, there are nine distinct files: DLOT, EEG, Eyetracker (comprising EYE and IVT data), GSR, NASA TLX (NSTLX), external events, log data generated from Psychopy (PSY), and affective data (TIVA). The multimodal dataset is quite extensive, covering various formats such as CSV, TXT, video, and XLSX, amounting to a total size of approximately 8 GB.

Additionally, the mental rotation task (MRT) images used in this study were a subset selected from the larger Peters & Battista dataset (https://www.uoguelph.ca/psychology/lab/mrt)[30,31], which contains a significantly larger collection of stimuli. This subset is not publicly available but can be accessed with permission from Dr. Michael Peters. All images used in this study have been compiled into a single document with structured metadata, including category number, image name, image, difficulty level, and match/no-match classification. Dr. Peters will provide access to this subset upon request. To obtain the exact images used in this study, researchers

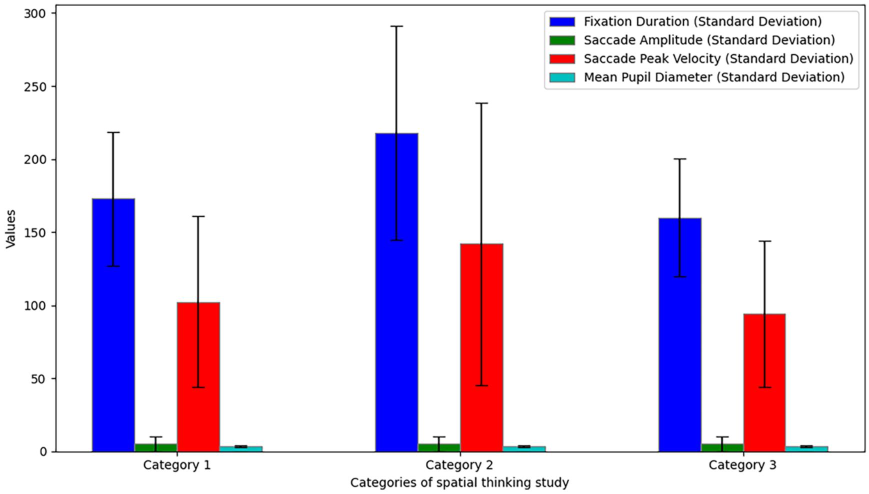

Fig. 4 Variation in eye-tracking measures across categories (error bars represent the standard deviation).

must reference this paper by providing the first author's name and the article title when contacting Dr. Peters at mpeters@uoguelph.ca for access. Alternatively, researchers may also contact the authors, and we will coordinate with Dr. Peters' team to facilitate access.

Technical Validation

Data acquisition. To ensure the quality and consistency of the collected data, participants were instructed to adhere to a standardized protocol under the supervision and guidance of the main researcher. Prior to commencing data collection, sensor calibration was performed for each participant to ensure accurate measurements. During the trials, participants were instructed to interact with the system while maintaining their natural body postures to minimize any potential bias in the generated movements.

Data synchronization. Data synchronisation was achieved by utilising built-in functions within the iMotions software. Additionally, the Unix system time stamp was employed to synchronize the data obtained from the Mind monitor with the data collected by iMotions. While we have implemented this synchronization protocol, it is important to note that it may not be entirely error-free. Therefore, we recommend users verify data synchronisation before utilizing it for analysis or research purposes.

Descriptive Analyses. We conducted various descriptive analyses to provide a comprehensive understanding of the collected data. These analyses include bar plots that offer an initial assessment of data quality and ensure that the data fall within plausible descriptive ranges. The following sections present visualizations for eye-tracking data, EEG frequency bands, GSR conductance, emotion recognition, and log data from PsychoPy.

Eye-Tracking Data. Figure 4 shows eye gaze features (Fixation Duration, Saccade Amplitude, Saccade Peak Velocity, and Mean Pupil Diameter) with standard deviations for the three categories of the spatial thinking study. Notably, 'Category 2' demonstrates the highest Fixation Duration, while 'Category 3' exhibits the most significant Saccade Amplitude. Moreover, 'Category 2' displays the highest Saccade Peak Velocity, signifying category variations. In contrast, the Mean Pupil Diameter remains relatively consistent across the categories. These insights into eye-tracking metrics can offer valuable avenues for further research and analysis within the study.

Electroencephalography. Figure 5 illustrates the average values of distinct EEG frequency bands: Delta, Alpha, Beta, Gamma, and Theta, accompanied by their standard deviations across the dataset comprising 38 participants. This graph provides valuable insights into the EEG frequency bands within various experimental categories. Notably, participants in 'Category 2' demonstrated notably higher Alpha band values, indicating potential cognitive distinctions. 'Category 3' participants exhibited elevated Beta band values, suggesting heightened cognitive engagement. Moreover, 'Category 2' participants displayed increased Theta band values, indicative of varying cognitive demands within this group. In the Gamma band, 'Category 2' exhibited higher values, implying differences in cognitive processing among categories. Lastly, 'Category 2' showcased elevated Delta band values, possibly indicating distinctions in cognitive engagement levels. The inclusion of standard deviations within each category further enriches our understanding of the variability in cognitive states. These variations in EEG patterns across categories hint at potential patterns or trends that researchers can explore in spatial thinking studies.

Galvanic Skin Response. Figure 6 illustrates the GSR (Galvanic Skin Response) Conductance CAL data for 38 individual participants over time. Each line on the plot corresponds to a different participant, labeled from 'Participant 1' to 'Participant 38'. The x-axis represents time, while the y-axis displays the GSR Conductance CAL value (Micro Siemens). As observed, the data patterns vary among participants, indicating unique physiological responses. This visualization offers a comprehensive view of how each participant's GSR Conductance values evolve over time, allowing for potential insights into their physiological responses during the experimental tasks.

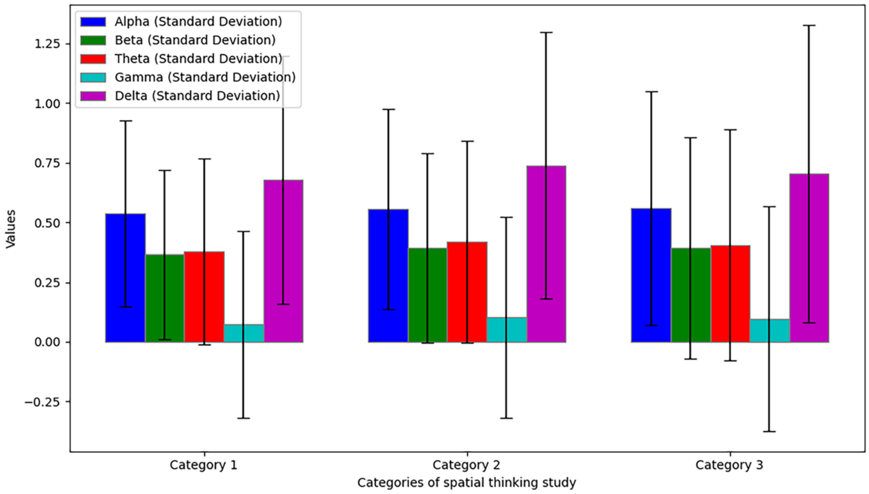

Fig. 5 Average EEG frequency band values across categories (error bars represent the standard deviation).

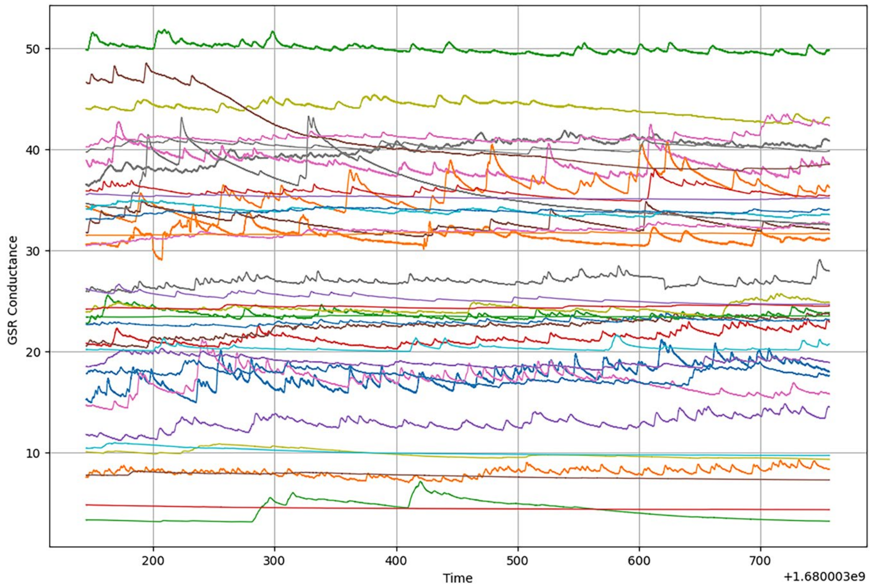

Fig. 6 GSR conductance CAL for 38 participants over time (error bars represent the standard deviation).

Emotion Recognition. Figure 7 showcases average counts and standard deviations for a sample set of emotions such as 'Engagement,' 'Surprise,' and 'Confusion' in different categories, shedding light on participants' emotional states during spatial thinking tasks. Also, Fig. 8 provides data on a sample set of facial metrics ('Lip Press,' 'Jaw Drop,' 'Brow Raise,' 'Chin Raise'), opening avenues for further investigation into participants' facial expressions and emotional responses. These differences in values with respect to categories present intriguing possibilities for future research directions.

Log Data from PsychoPy. A sample visualization using the log data is shown below. The average response time was calculated from this log data for all the participants in the three categories. The results are presented in Fig. 9, which shows the trends in average response time with respect to each category. It was observed that the average response time reduced over the course of the experiment. Additionally, the data showed that the average response time was slightly higher for males compared to females in category 1. In Fig. 10, we present another sample image that displays the number of correct answers of male and female participants with respect to the three different categories. Figures 9 and 10 provide valuable insights into the participants' performance in the spatial thinking tasks, highlighting the temporal patterns of response times and potential gender-related differences in task performance. These observations contribute to a better understanding of the spatial thinking process in postgraduate students and the implications of gender-related factors on task performance.

To ensure dataset reliability, reaction times (RTs) and accuracy were analyzed across different difficulty levels. As shown in Fig. 11, RTs increased from the easy to the medium level but plateaued or even decreased at the

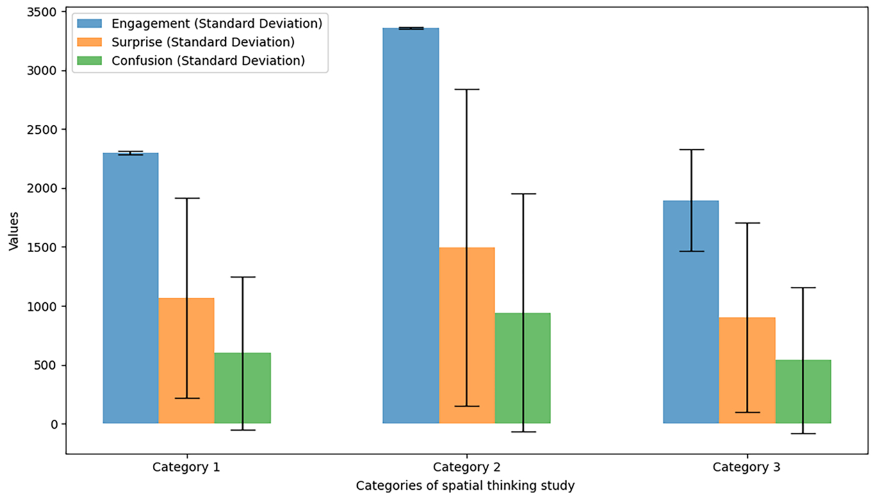

Fig. 7 Variation in emotion counts across categories (error bars represent the standard deviation).

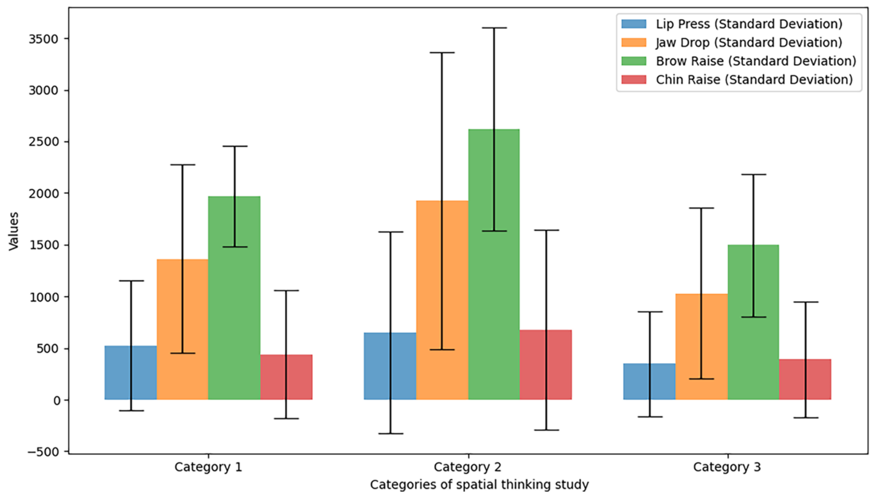

Fig. 8 Variation in action unit counts across categories (error bars represent the standard deviation).

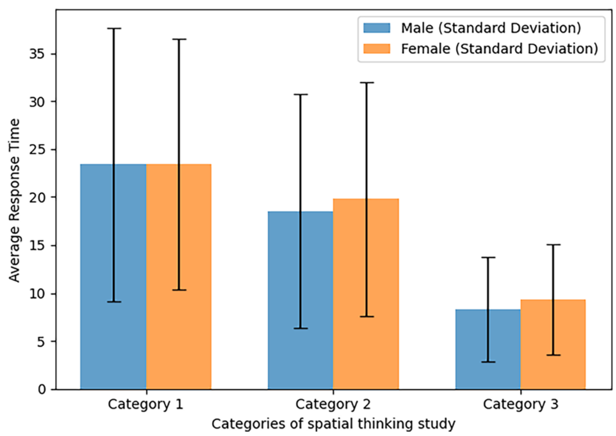

Fig. 9 Average response time by category and gender (error bars represent the standard deviation).

difficult level, likely because participants moved on to subsequent questions rather than spending additional time. Figure 12 illustrates a corresponding decrease in accuracy as difficulty increased. Together, these analyses

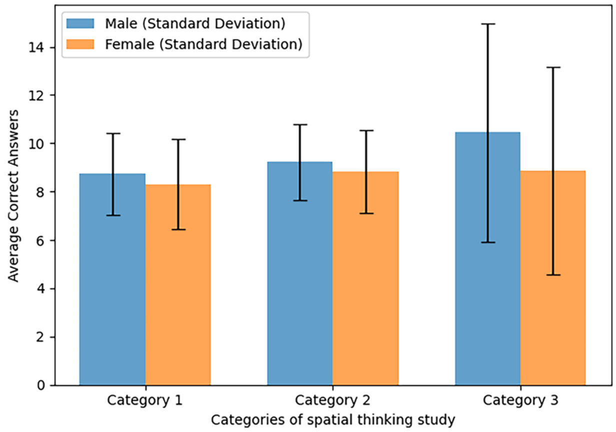

Fig. 10 Average correct answers by category and gender (error bars represent the standard deviation).

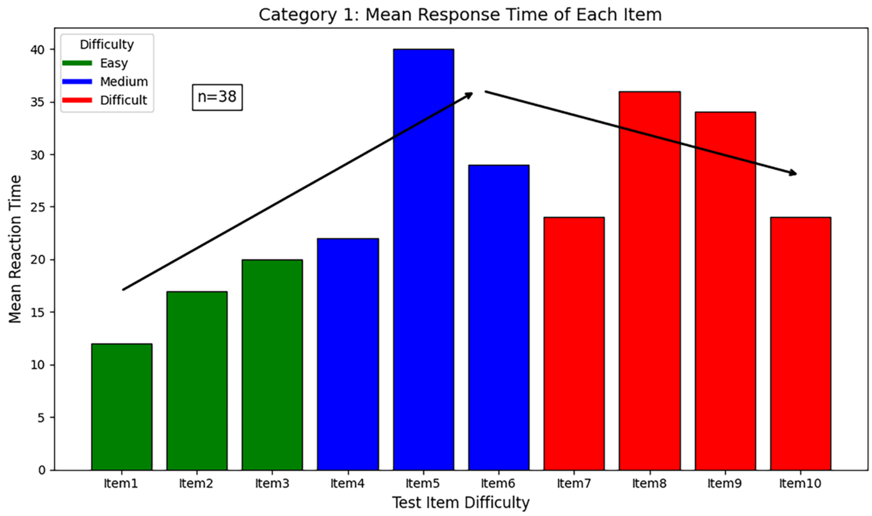

Fig. 11 Validating the Test Design: Reaction Time vs Item Difficulty.

confirm the robustness of the questionnaire for studying mental rotation performance under varying difficulty conditions.

Usage Notes

In order to facilitate the replication of our study design and sensor data collection, we provide detailed information about the file structure and export options. The data file used in the study has the extension ".csv" and is separated by commas. It encompasses various sections and metadata, contributing to a comprehensive understanding of the study. The metadata and detailed explanations are available in the appendix section of this document (Refer to Appendix Sections 2 to 5 for detailed metadata pertaining to the data collected from sensors). The database contains only CSV files, each with a column header corresponding to the attribute names. Additionally, the information or explanations of the attributes are provided solely in the appendix section of this document.

For data export, the file offers various options. The "Sensor Data" export provides a comprehensive raw data export with synchronized data across multiple modalities. Users can select specific respondents, stimuli, and timestamp formats. The exported file is tab-separated and compatible with software such as Microsoft Excel. The "Exposure Summaries" export generates a file listing all stimuli exposures, including respondent names, stimulus names, actual exposure times, and order of presentation. This exposure summary is available in the database for each participant and is saved as a text file with the name 'exposure_summary.'

Study settings, including the desktop study environment, screen resolution of 1920 × 1080, and default calibration mode, are detailed in this section. The gaze analysis settings, such as the I-VT fixation filter, discarding short fixations, noise reduction using the moving average method, and merging adjacent fixations, are also specified. Live markers utilized in the study include the default marker named "Default Marker" with no modifier, assigned to the F3 key and represented by the color orange. These settings serve as a guide for replication purposes. For comprehensive details, please refer to Section 6 of the Appendix. We also have a website (https://sites.google.com/view/etiitb-multimodal-database) that provides guidelines for downloading and utilizing the data. This website redirects to Figshare for downloading the dataset.

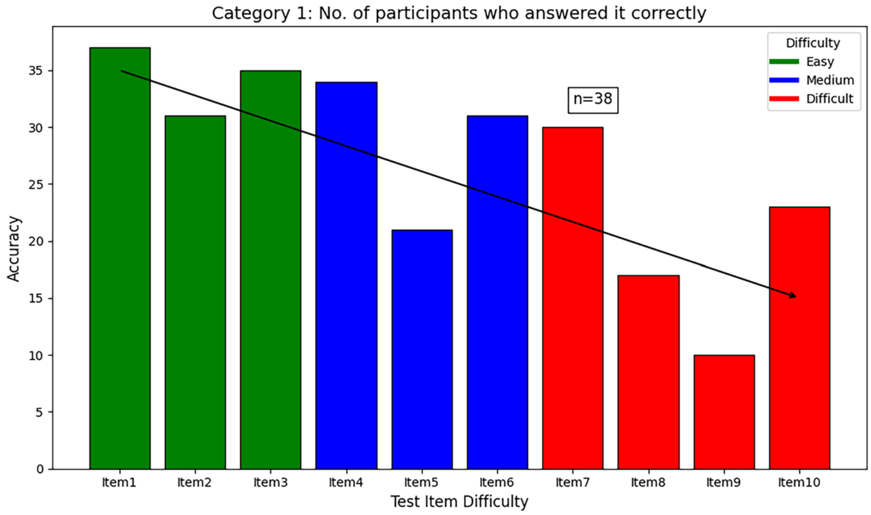

Fig. 12 Validating the Test Design: Accuracy vs Item Difficulty.

Code availability

We provide the raw CSV, TXT, and XLS files obtained during the data collection. No custom code was implemented for generating or processing the data. The entire dataset is anonymized. The alignment of iMotions data with the Mind monitor data was performed using the systems' Unix timestamp.

Received: 10 January 2024; Accepted: 20 March 2025;
Published online: 03 April 2025

References

1. Freedman, R. J. & Rovegno, L. Ocular dominance, cognitive strategy, and sex differences in spatial ability. *Perceptual and Motor Skills* **52**, 651–654 (1981).
2. Gilger, J. W. & Ho, H. Z. Gender differences in adult spatial information processing: Their relationship to pubertal timing, adolescent activities, and sex-typing of personality. *Cognitive Development* **4**, 197–214 (1989).
3. Hegarty, M. Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review* **25**, 1212–1219 (2018).
4. Moen, K. C. *et al.* Strengthening spatial reasoning: Elucidating the attentional and neural mechanisms associated with mental rotation skill development. *Cognitive research: principles and implications* **5**, 1–23 (2020).
5. Maeda, Y. & Yoon, S. Y. A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: Visualization of rotations (psvt: R). *Educational Psychology Review* **25**, 69–94 (2013).
6. Voyer, D. Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review* **18**, 267–277 (2011).
7. Griksiene, R., Monciunskaite, R., Arnatkeviciute, A. & Ruksenas, O. Does the use of hormonal contraceptives affect the mental rotation performance? *Hormones and Behavior* **100**, 29–38 (2018).
8. Heil, M., Kavšek, M., Rolke, B., Beste, C. & Jansen, P. Mental rotation in female fraternal twins: Evidence for intra-uterine hormone transfer? *Biological Psychology* **86**, 90–93 (2011).
9. Lourenco, S. F. & Liu, Y. The impacts of anxiety and motivation on spatial performance: Implications for gender differences in mental rotation and navigation. *Current Directions in Psychological Science* **09637214231153072** (2023).
10. Khooshabeh, P., Hegarty, M. & Shipley, T. F. Individual differences in mental rotation. *Experimental Psychology* **60**, 164–171 (2013).
11. De Lisi, R. & Cammarano, D. M. Computer experience and gender differences in undergraduate mental rotation performance. *Computers in Human Behavior* **12**, 351–361 (1996).
12. Moè, A. Does experience with spatial school subjects favour girls' mental rotation performance? *Learning and Individual Differences* **47**, 11–16 (2016).
13. Guay, R. B. & McDaniel, E. D. The relationship between spatial abilities and social-emotional factors [microform]. https://nla.gov.au/nla.cat-vn5420782 Accessed: 08 August 2024 (1982)
14. Bartlett, K. A. & Camba, J. D. Gender differences in spatial ability: A critical review. *Educational Psychology Review* **35**, 8 (2023).
15. Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. *Science* **171**, 701–703 (1971).
16. Just, M. A. & Carpenter, P. A. The role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation* **8**, 139–143 (1976).
17. Heil, M. & Jansen-Osmann, P. Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology* **61**, 683–689 (2008).
18. Nazareth, A., Killick, R., Dick, A. S. & Pruden, S. M. Strategy selection versus flexibility: Using eye-trackers to investigate strategy use during mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **45**, 232 (2019).
19. Hsing, H.-W., Bairaktarova, D. & Lau, N. Using eye gaze to reveal cognitive processes and strategies of engineering students when solving spatial rotation and mental cutting tasks. *Journal of Engineering Education* **112**, 125–146 (2023).
20. Xue, J. *et al.* Uncovering the cognitive processes underlying mental rotation: An eye-movement study. *Scientific Reports* **7**, 10076 (2017).
21. Gardony, A. L., Eddy, M. D., Brunyé, T. T. & Taylor, H. A. Cognitive strategies in the mental rotation task revealed by eeg spectral power. *Brain and Cognition* **118**, 1–18 (2017).
22. Ding, X., Guo, X., Lo, T. T. S. & Wang, K. The spatial environment affects human emotion perception-using physiological signal modes. *CAADRIA proceedings* https://api.semanticscholar.org/CorpusID:250969878 (2022).
23. Ruotolo, F., Claessen, M. & Van der Ham, I. Putting emotions in routes: The influence of emotionally laden landmarks on spatial memory. *Psychological research* **83**, 1083–1095 (2019).
24. Ashwin, T. S., Shafi, S. D. & Ramkumar, R. Dlot: An open-source application to assist human observers. In *31st International Conference on Computers in Education*, vol. 1, 478–483 (Asia-Pacific Society for Computers in Education (APSCE), 2023).

25. Lyons, I. M. *et al*. Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical cognition* **4**, 526–553 (2018).
26. Gosling, S. D., Rentfrow, P. J. & Swann Jr, W. B. A very brief measure of the big-five personality domains. *Journal of Research in personality* **37**, 504–528 (2003).
27. Pintrich, P. R. & De Groot, E. V. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology* **82**, 33 (1990).
28. Hart, S. G. & Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, 139–183 (Elsevier, 1988).
29. F Caissie, A., Vigneau, F. & A Bors, D. What does the mental rotation test measure? an analysis of item difficulty and item characteristics. *The Open Psychology Journal* **2** (2009).
30. Peters, M. & Battista, C. Applications of mental rotation figures of the shepard and metzler type and description of a mental rotation stimulus library. *Brain and Cognition* **66**, 260–264 (2008).
31. Peters, M. *et al*. A redrawn vandenberg and kuse mental rotations test-different versions and factors that affect performance. *Brain and cognition* **28**, 39–58 (1995).
32. Ganis, G. & Kievit, R. A. A new set of three-dimensional shapes for investigating mental rotation processes: Validation data and stimulus set. *Journal of Open Psychology Data* (2015).
33. Peters, M. Sex differences and the factor of time in solving vandenberg and kuse mental rotation problems. *Brain and cognition* **57**, 176–184 (2005).
34. Isaacs, E. B. & Vargha-Khadem, F. Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology* **7**, 377–380 (1989).
35. Kane, M. J., Conway, A. R., Miura, T. K. & Colflesh, G. J. Working memory, attention control, and the n-back task: a question of construct validity. *Journal of Experimental psychology: learning, memory, and cognition* **33**, 615 (2007).
36. Asif, A., Majid, M. & Anwar, S. M. Human stress classification using eeg signals in response to music tracks. *Computers in Biology and Medicine* **107**, 182–196 (2019).
37. Pekrun, R. & Linnenbrink-Garcia, L. Academic emotions and student engagement. In *Handbook of research on student engagement*, 259–282 (Springer, 2012).
38. Calvo, R. A. & D'Mello, S. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* **1**, 18–37 (2010).
39. T S, A. & Biswas, G. Relating students cognitive processes and learner-centered emotions: An advanced deep learning approach. In *Proceedings of the 26th International Conference on Multimodal Interaction*, ICMI '24, 575–584, https://doi.org/10.1145/3678957.3685751 (Association for Computing Machinery, New York, NY, USA, 2024).
40. Gupta, A., D'Cunha, A., Awasthi, K. & Balasubramanian, V. Daisee: Towards user engagement recognition in the wild. *arXiv preprint arXiv:1609.01885* (2016).
41. Ashwin, T. S. *et al*. A multisensor dataset of south asian postgraduate students working on mental rotation tasks https://doi.org/10.6084/m9.figshare.28120670 (2025).

Acknowledgements

This work was carried out through seed funds provided to Prof. Ritayan Mitra by IIT Bombay (IRD/0517-IRCCSH0-001) and the project fund of Prof. Ramkumar Rajendran (SRG/2020/000279).

Author contributions

Ritayan Mitra, Karishma Khan, Kabyashree Kanikar, Ramkumar Rajendran, and Ashwin T S designed the project. Kabyashree Kanikar, Karishma Khan, and Suraj Ranganath collected the data. Suraj Ranganath was responsible for data anonymization and validation. Ashwin T S, Karishma Khan, and Suraj Ranganath performed the data analysis. Ritayan Mitra and Kabyashree Kanikar managed the project. Ashwin T S wrote the main manuscript, with editing by Ashwin T S, Ritayan Mitra, and Ramkumar Rajendran. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04865-5.

Correspondence and requests for materials should be addressed to A.T., R.R. or R.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.