

Systems Thinking Ability Instrument Description

The study will be split into two sections. In the first section, participants will complete the assessment game, which entails exploring the various systems around the space ship and completing the five measures related to systems thinking ability (Awareness of System Elements, Identification of System Relationships, Understanding of System Dynamics, Evaluate and Revise, Application to the Mission) as well as measures of Creativity, Curiosity, and Openness to Information. Participants will have up to four hours to complete this section of the study. The second section will assess participants on hierarchical working memory, cognitive complexity, cognitive flexibility, pattern recognition, and spatial ability. Participants will have up to two hours to complete this section of the study.

Participants will complete either one or both assessment sections, based on their availability.

Section 1:

In the first section, participants complete the assessment game, which entails exploring the various systems on the space ship and completing assessments for five dimensions of systems thinking ability (Awareness of System Elements, Identification of System Relationships, Understanding of System Dynamics, Evaluate and Revise, Application to the Mission) as well as measures of Creativity, Curiosity, and Openness to Information. The length of the assessment will vary based on participant choices, but they will have up to four hours to complete this section.

The participant communicates his/her findings with Earth forces through a special Mission Console and receives assignments from them to explore various aspects of the ship and its personnel. In the Mission Console, there are files that store information about each system being explored. The participant updates the files with information he or she learns about the systems, then answers questions about the systems.

Participants focus on one subdimension at a time for any given system, starting with *Awareness of Elements*, but can work on multiple systems at the same time. Once they believe they have identified all of the elements of a given system, they submit their work and the system checks their responses and provides feedback. This ensures all test takers start on a level playing field for the next subdimension (e.g., Identifying Relationships).

After completing Awareness of Elements and Identifying Relationships, in *Understanding System Dynamics*, the test taker is asked to “calibrate” software that will ostensibly analyze the information provided about the system and send that information to Earth. During this calibration, the test taker will be answering questions of increasing complexity about system archetypes and complex system dynamics—the subdimension. Next, the test taker will be told that there may be new information available, so they will be prompted to review each system. In *Evaluate and Revise* the test taker will have the ability to update their conceptualizations based on a review of each system. Lastly, the test taker must demonstrate understanding of the implications of system relationships and dynamics by answering multiple-choice questions in the *Application to Mission* subdimension. Scoring will be based on the number of correct responses for each dimension, or points acquired by comparing the participant responses to expert responses.

Awareness of System Elements: The test taker is presented with a graphic list of the observations they collected for a given system. From these observations, the test taker sorts the components of the system into appropriate subsystems while avoiding distractor items. As an example, the test taker will identify the following elements as part of the automatic elevator

operation subsystem: Mode Switch, Elevator Call Room Console, Elevator Car, Sentinel Key, Destination Map, Drive Stick, Elevator Doors, Release/Catch button. After the test taker inputs enough elements, he/she will be able to submit his/her responses for feedback. The feedback will show which elements were omitted and which were included in error. This will put each test taker on a level playing field going into the next module. Responses are compared to known answers or expert ratings.

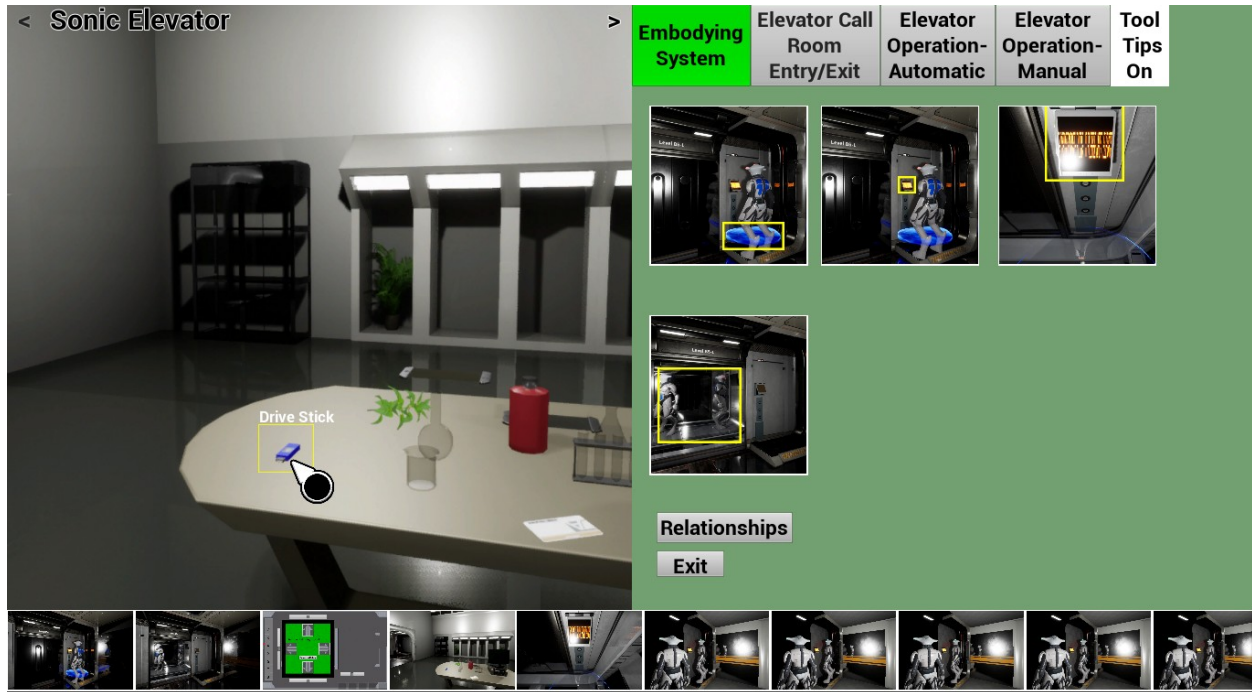


Figure 1: Awareness of Elements Interface

Identification of System Relationships: This module asks the test taker to generate causal statements to describe the dyadic relationships between important elements in the system. After the Awareness of Elements module is complete, labels will be added to the images in the elements menu so that the test taker can generate statements about the elements' relationships with each other. In Identifying Relationships, the test taker will be presented with three side-by-side drop-down menus for each statement. Each statement will be structured as follows: [Element 1] [Relational verb / verb phrase] [Element 2]. The drop-down menu for each element will consist of the full list of elements relevant to the system. The test taker will also select a verb that best fits the relationship between the two selected elements (e.g., increases, decreases, feeds into, is fed by, etc., as appropriate). Selecting two elements and a causal verb between the two elements creates a Causal Sentence. Examples of system relationships in the Automatic Elevator Operation subsystem include:

- Selecting Automatic Mode *activates* a request for authentication
- Selecting a destination *activates* an Elevator Car
- Sentinel Key *activates* Destination Map
- Sentinel Key *deactivates* a request for authentication
- Drive stick *activates* Destination Map
- Drive stick *deactivates* a request for authentication

- Release/Catch button *activates* Elevator Car
- Release/Catch button *deactivates* Elevator Car

Test taker scores are based on total correct relationships. Responses will be compared to correct responses if these are known, or expert ratings if objectively correct answers are not obtainable. After the test taker submits their statements, they will be given feedback in the form of the correct causal statements. The test taker must complete the Identifying Relationships module and commit answers before receiving feedback and moving to the next module.

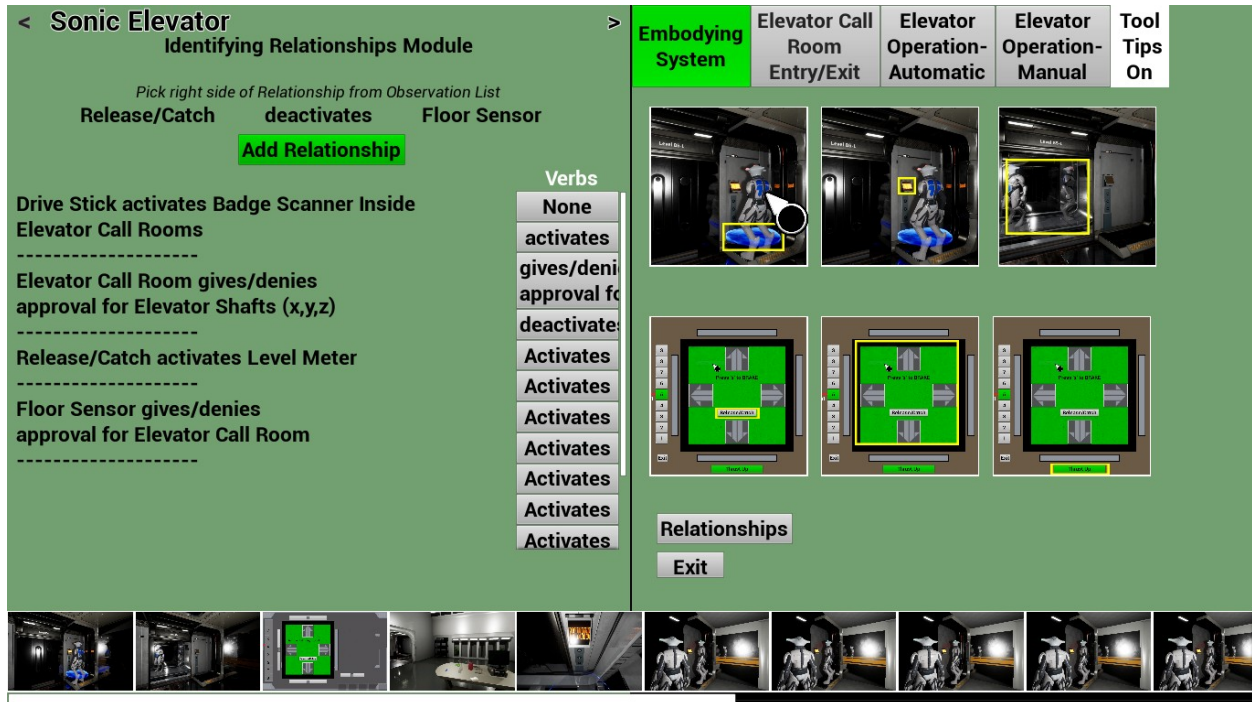


Figure 2: Identifying System Relationships Interface

Understanding of System Dynamics: Understanding System Dynamics is measured by first providing the basic language for describing feedback loops, as well as providing a basic familiarity with system archetypes (e.g., see Figure 1). During this introduction to system dynamics, the test taker is shown simple feedback loops and system archetypes and asked to categorize them.

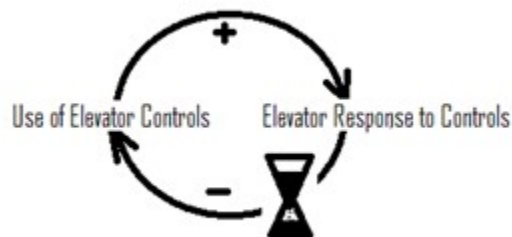


Figure 3: Elevator System Archetype Diagram

This process is completed three to four times. For each system, the questions posed to the test taker get harder by reflecting more complex system archetypes. A sample question is:

The Mode Switch determines the operation type of the elevator: manual or automatic. When set to manual, the elevator doors open and the elevator control panel is activated. When set to

automatic, the call room console is activated and the elevator control panel is deactivated. Is the relationship between the automatic and manual mode a feedback loop?

- a. Yes, Reinforcing loop
- b. Yes, Balancing loop
- c. Yes, combination of loops
- d. No, not a feedback loop

Finally, the user needs to choose the systems diagram that best depicts the system's dynamics, and receive feedback about the category that is chosen. The test taker's score on Understanding System Dynamics is determined by his or her performance categorizing systems in the calibration mode.

Application to the Mission: Application to the Mission will be assessed with multiple choice questions about the systems that are explored in the Mission Console. After the test taker completes the Evaluate and Revise module, the Application to the Mission module is initiated. In this module, the test taker answers questions about the relevance of the system, its structure, its similarity to other systems, logical next actions to take to influence and control the system, etc. An example item is:

If you were able to deactivate all of the Swarmbot zones (i.e., the areas generated by Location Balancers), which of the following would be the most likely possibility?

- a. You would have completely unrestricted access to all parts of the ship.
- b. The Swarmbots would swarm uncontrollably throughout the ship.
- c. The exterior of the ship would be unguarded.
- d. The Swarmbots would leave the ship.

Scores are based on similarity to expert responses.

Participants in Section 1 will also complete measures of curiosity, openness to information, and creativity.

Curiosity will be measured by allowing test takers to investigate novel objects. As part of the game, subjects will be placed in a room full of alien artifacts where they must wait for a predetermined period with a distractor task, but no mission-related activities (see Figure 2).



Figure 4: Room of display objects

By making the test taker wait, we ensure that the subject does not have to choose between a mission task and exploring, which could measure decision-making rather than just curiosity. A

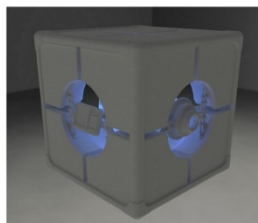
set of 30 items will be available in the room that participants can click on multiple times to learn more and more about the object (see Figure 3 for example levels). Scoring will be based on length of exploring and depth of exploring, which will be measured by how many levels of information that test taker pursues for each object. Distractor tasks will be included so the test taker has the option to do a familiar task instead of investigate their surroundings.

Durag's Puzzle (Popular Game)



1. Origin: Saxum star in the Farose galaxy
2. The puzzle is solved when the textures, patterns, and height of each cube are aligned correctly.
3. Durag's puzzle is approximately 9x9x9 inches with 2,359,085 possible combinations.
4. The fastest known time for solving Durag's Puzzle is 23 centiseconds.

Zaplink (Communication Device)



1. Origin: Jkgatshri planet, created by the Driehurg species
2. Purpose: Communication unit in which gamma rays are transmitted across galaxies, destroying organic material that intercepts the message
3. The unit is made of rock from the Jkgatshri planet, which produces and absorbs the gamma rays.
4. Gamma rays have the smallest wavelengths and the most energy of any wave in the electromagnetic spectrum.

Figure 5: Sample Curiosity objects that test taker can investigate

Openness to information will be measured with a series of 40 questions that assess the test taker's understanding of the ship based on initial assumptions, observations and then subsequent planted information that reveals an alternative reality. For example, the opening statements refer to the ship being a warship and how U.S. aircraft and service members are disappearing. The Sentinel robots look and act like a military force. In reality, the ship is a business enterprise with Alien X as the CEO. Their focus is conducting research to develop various medicines, such as one to extend the lifespan of certain species. The test taker rates the likelihood of statements, such as "*The overall mission of the alien spacecraft is to:* 1. Assess the capabilities of various planets in order to plan future attacks 2. Search for new technology to acquire 2. Invade and take over planets..." or "*The individual or group that owns the spacecraft is probably:* 1. A band of alien criminals, 2. A wealthy alien individual 3. An alien government..."

To measure **creativity**, the test taker will be presented with a situation in which he/she needs to generate multiple ways to collect water in a confined room with water sources. The test taker is instructed: "Use the objects in the room to design as many ways as possible to fill the container with water from one of the sources of water in this room. For each attempt, use only one source of water to fill your container. You have five minutes to come up with as many solutions as possible. You collect points for each design you submit." Objects in the room will include items such as a sponge, a fork, a dishtowel, a straw, books, etc. The test taker's creativity score will be based on the number of solutions they generate, the originality of the solutions, and the feasibility of the solutions.

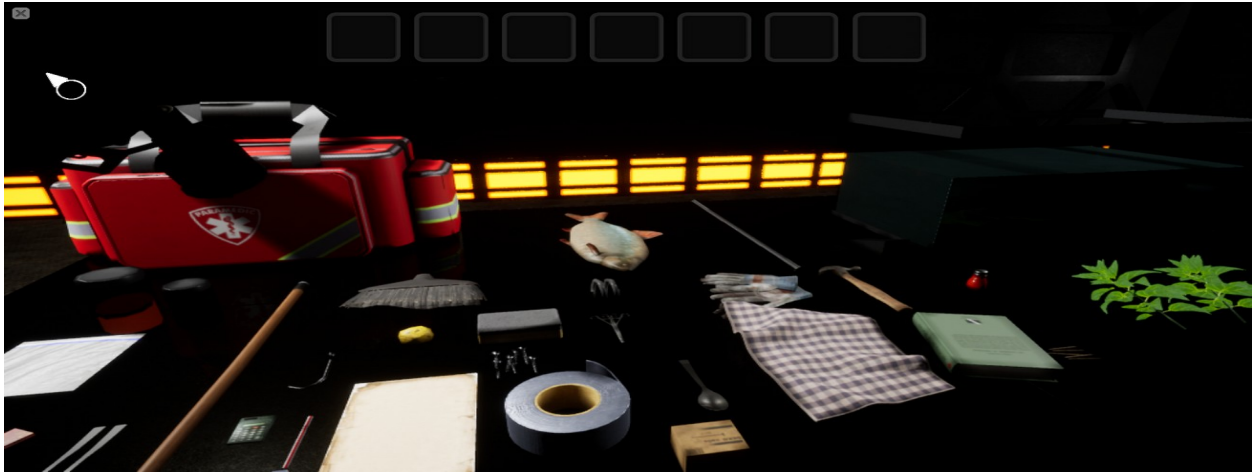


Figure 6. Creativity screenshot

Section 2:

In the second section, participants will complete the Phase 1 construct measures: Hierarchical working memory capacity, Cognitive flexibility, Cognitive complexity, Pattern Recognition (includes two parts: Anomaly Detection and Multi-level figure recognition), Spatial Ability (includes “Find the shapes” (intrinsic/static), “Shape cross-sections” (intrinsic/dynamic), “Visual estimation/water jar” (extrinsic/static), and “Spatial perspective-taking” (extrinsic/dynamic)). In the future these measures will be a component of the assessment game, but they are currently separate to reduce testing time. Participants will have up to two hours to complete this section.

These measures are described in more detail below:

Hierarchical Working Memory

The Hierarchical Working Memory measure presents a sequence of location-number pairs that participants have to remember. The pairs are tested in a random order (not in order of presentation). The test uses a small set of locations and numbers repeatedly throughout all trials, so that across successive trials the same locations and numbers are paired in many different ways. Accuracy, therefore, depended on remembering the pairing in the current trial as opposed to different pairings from previous trials. In the first set of items, locations are in one of nine boxes in a 3 by 3 grid and one of ten digits are used. During the assessment, subjects are presented with a series of location-number pairs then are asked to recall those location-number pairs.

In the second set of items that are presented, the same approach will be used but the locations will be presented as being in different levels of a hierarchy. For example, zooming between a nucleus, an atom, and a molecule would show three levels of a hierarchy, each of which could house locations to store numbers. Another example is spots on a cow, the full cow, and the field that cows graze in (see Figure 4). Subjects will have to recall the number and location, where the location is both the correct level and the correct place in that level. This measure should take no more than 5 minutes to complete.

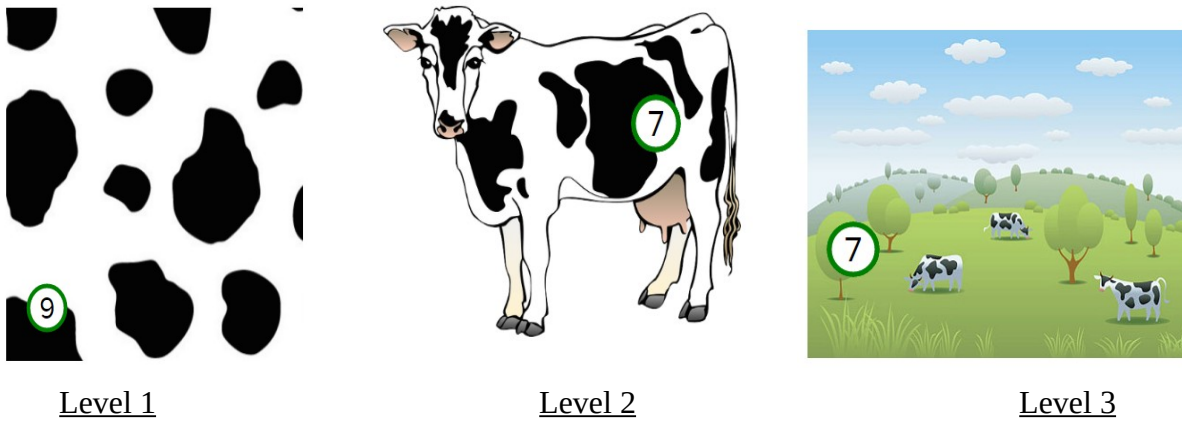


Figure 7: levels of hierarchical working memory task

Cognitive Complexity

The Cognitive Complexity measure is based on an object-sorting task. Subjects are presented with a list of objects and asked to group the objects based on commonalities (see Figure 5). Beginning with group 1, subjects identify one commonality, label the group, and put in as many objects as there are sharing that characteristic. When group 1 is complete, the subject starts group 2 and repeats the sorting task with the initial list of objects.

Create a group by clicking all the objects with similarities in **how they look**. Then, give the group a descriptive label by clicking where it says "Name this group" and type in a label. Click "Save" when you're ready to create another group. You can use the objects over in as many groups as you like. You can edit or delete a group by clicking on the icons to the right of each group in the list.

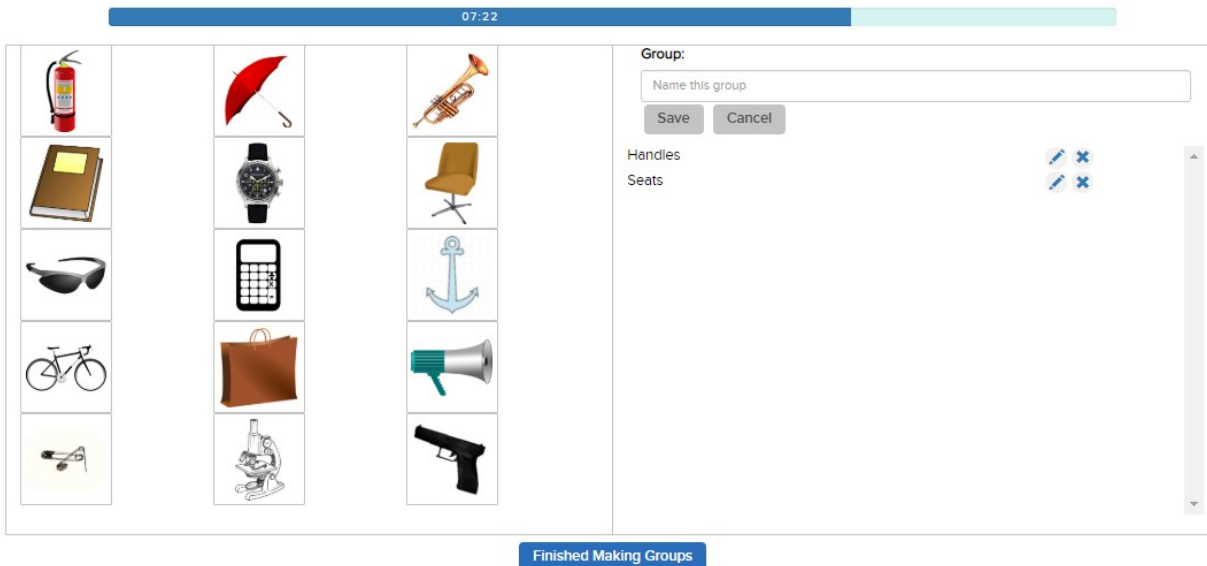


Figure 8: Screenshot of Cognitive Complexity object grouping task

Subjects are told to continue in this fashion, creating as many groups as they can. The test uses verbal and pictorial stimuli in separate rounds. The verbal stimuli are the names of common systems. Test-takers interact with stimuli by dragging and dropping them into groups and labeling the groups. This measure should take no more than 27 minutes to complete.

Cognitive Flexibility

Cognitive flexibility is measured with a scenario-based hypothesis formation task (see Figure 6). The test taker is presented with a short description of a scenario involving an unexplained event

or phenomenon and asked to select hypotheses about these events or phenomena. The test taker is also asked to rank order the five hypotheses from most likely to least likely.

Figure 9: Screenshot of Cognitive Flexibility scenario and Guesses

Once the test takers have selected and ranked five hypothesis for why the first scenario is occurring, they are given additional information about the scenario and asked to add to, modify, or delete their original hypotheses. This process is repeated four times with the test taker receiving additional information and being asked to make changes to his/her hypotheses and the hypotheses' rankings. This module should take no more than 15 minutes to complete.

Pattern Recognition

The Pattern Recognition measure that we developed has two phases. In the first phase, respondents will watch a factory-like setting, tracking inputs and outputs on a conveyer belt, with the goal of identifying when behavior begins to deviate or when abnormal activity occurs (see Figure 7). At different points along the conveyer belt, there will be input funnels or stations where additional inputs will be added to the product. The rules for creating a product on the assembly line will generally be consistent throughout the trial, however, there will be occasional deviations. The test-taker's task will be to monitor the assembly line and determine when deviations occur.

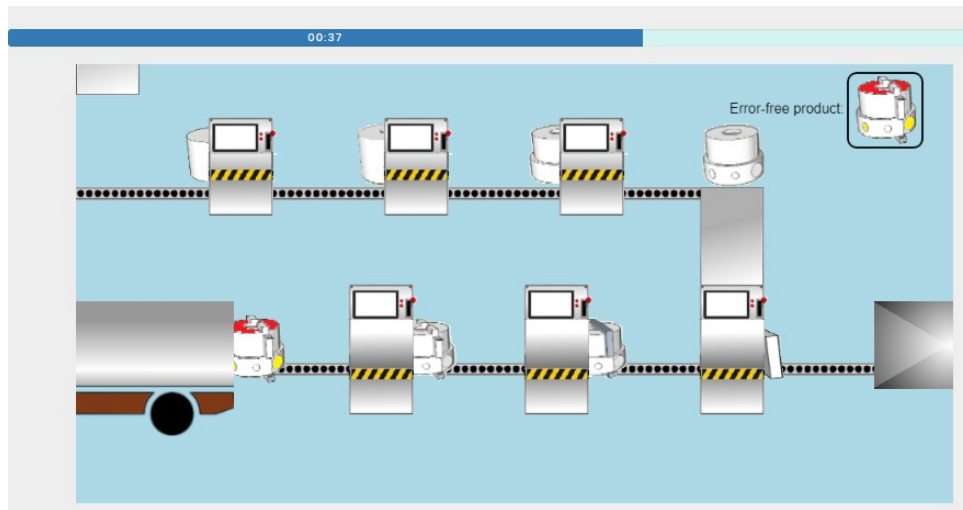


Figure 10: Screenshot of Factory Assembly line in the Pattern Recognition task

In the second phase of the Pattern Recognition measure, the test-taker is given a short time to examine a multi-level Navon figure, then the figure disappears and the test-taker must report the letters or numbers that they saw in the figure from smallest to largest (see Figure 8). This phase also asks test-takers to respond quickly to just one level – multi-level figures are presented for a short interval and the user states what letter or number he/she saw at each a micro, meso, and macro level.

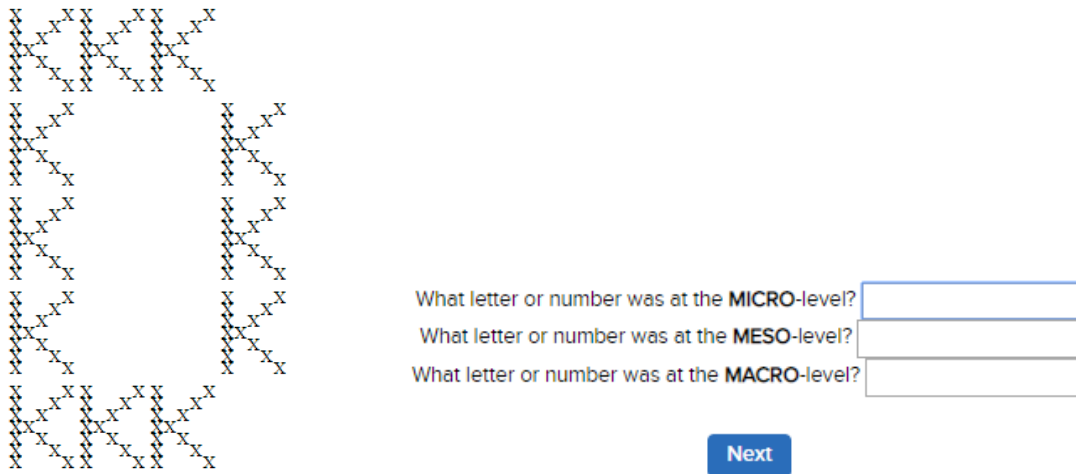


Figure 11: Multi-level Navon figure and Follow up questions in the Pattern Recognition task

Scores on this task depend on the user's accuracy at each level. Participants' reaction times are measured for responses at each level. This module should take no more than 15 minutes to complete.

Spatial Ability

The Spatial Ability measure we developed focuses on four types of spatial reasoning: intrinsic-static, intrinsic-dynamic, extrinsic-static, extrinsic-dynamic. There are specific items for each type of spatial reasoning.

Type 1: Test-takers are presented with a simple shape (e.g., an octagon; a box) and told to study that shape and look for it in the next image, which is more complex (see Figure 9). The complex image appears and the simple image disappears. The complex image is drawn on the inner

surface of a perceived tube or tunnel in an arch from left to right. The participant perceptually moves through the tunnel and is prompted to find as many of the simple shapes in the complex images as possible within a set period of time. After that period of time elapses, or after all the simple images have been found, the next item is presented. In the next item, a new simple shape and new section of the tunnel are used.

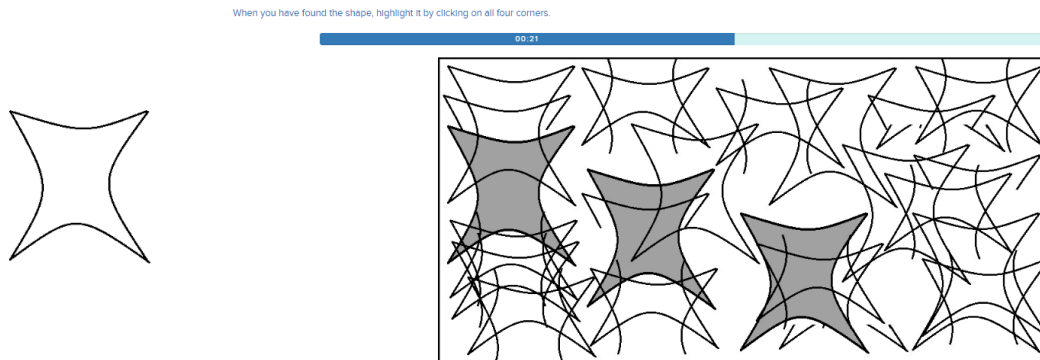


Figure 12: Spatial Ability task where a simple shape must be found and highlighted in the more complex image

Type 2: The test-taker is presented with a three dimensional shape with a plane drawn through it and asked to imagine the shape was cut along that plane (see Figure 10). The subject must identify the two-dimensional shape that would result as the surface cut by this plane. The correct response is chosen from among five options.

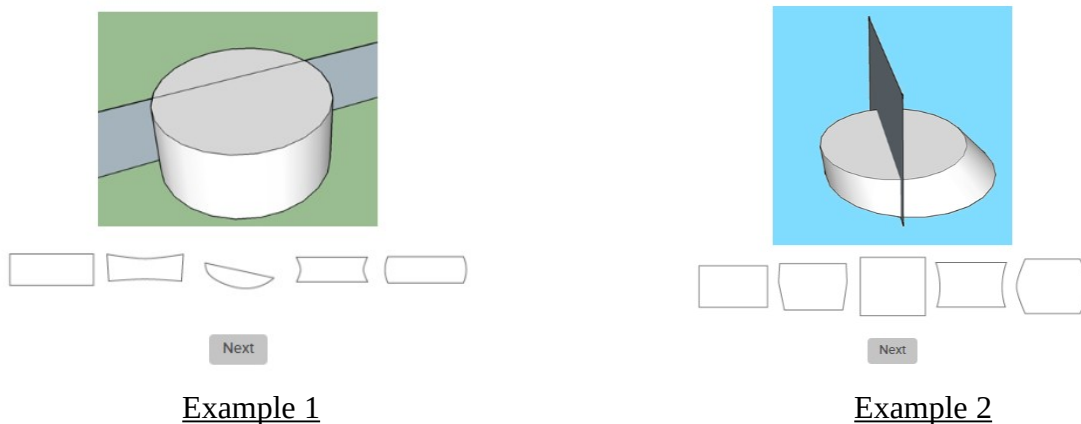


Figure 13: Spatial Ability task where a 2D shape must be determined based on the plane of a 3D shape

Type 3: The test-taker is presented with a series of transparent containers that are tilted at angles varying from +/- 1 to +/-90 (see Figure 11). Each bottle has an amount of fluid in it marked by horizontal line. The task is to indicate the height that the fluid line would reach if the bottle were upright.

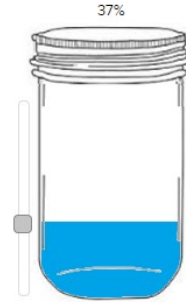


Figure 14: Spatial Ability task where test taker must determine the fluid level of a tilted Mason jar

Type 4: The test taker is shown a picture of a scene from one perspective, with major objects in the view lettered A through E (see Figure 12). Then a second picture with a different perspective is shown, with major objects overlapping from the first picture.



Figure 15: Spatial Ability task where images from different perspectives are shown

The test-taker is asked which letter corresponds to the position of someone who would have the perspective depicted in the second picture. Once a letter is selected, the test-taker must draw the angle of orientation required to obtain the view shown in the second picture (see Figure 13). Combined, these modules should take no more than 50 minutes to complete.

Where in Picture 1 would someone need to stand to take Picture 2?

What direction would the camera need to be pointing to produce Picture 2?
Please choose the answer that is closest.



Figure 16: Spatial Ability task where test taker must determine the angle and perspective of the previous images

Demographics and Feedback

All participants will be asked to provide basic demographic information (age, gender, level of school achieved, experience with video games), as well as feedback on aspects of their experience with the assessments, including an opportunity to provide written comments and suggestions. Section 1 participants will be asked to provide feedback on game functionality and their experience with the game, and Section 2 participants will be asked to provide feedback on the functioning of the Phase I assessments. The demographics and feedback measures should take no more than 5 minutes to complete.