Overview

Welcome to the Human Evaluation Survey!
Before start to the survey, please read the instruction below.

Instructions:

❖ We are going to show you two different Sequences. Both Sequences consist Textual and Video Instructions. Given the <mark>\<Task\></mark> (for example, Task: How to make pancakes);

➢ Compare Two Sequences of Steps (**Sequence 1** and **Sequence 2**).

➢ **Sequence 1 Videos were generated by using \<text\> and \<context\> instructions**

➢ **Sequence 2 Videos were generated by using \<visual\> instructions**

➢ Choose which Sequence is better in terms of **4 aspects**:

1. **Textual Informativeness:** whether the Textual Sequence (*sequence of texts*) contain the amount of information needed to complete the task.

2. **Visual Informativeness:** whether the Visual Sequence (*sequence of videos*) contain the amount of information needed to complete the task.

3. **Temporal Coherence:** whether the Multimodal sequence (*paired sequence of texts and videos*) meet the temporal common-sense requirements, such as a step occurs before another step instead of after.

4. **Plan Accuracy:** whether the Multimodal sequence (*paired sequence of texts and videos*) can successfully complete the task.

What's the step-by-step procedure for \<task\>?
Task: How to Make a Gender Reveal Balloon?

Sequence 1:

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8
Text: "Materials needed" **Context:** "Gender reveal balloons (you can buy these online or at party supply stores), balloon pump, ribbon or string, tape (optional)."	**Text:** "Inflate the Balloons" **Context:** "Inflate the gender reveal balloons with air using a balloon pump. Make sure to inflate them to the desired size, but not too tightly, as you want them to be easy to pop when needed."	**Text:** "Attach the Ribbon or String" **Context:** "Attach a ribbon or string to each balloon. This will make it easier to hold onto the balloons and pull them apart when it is time for the reveal."	**Text:** "Add the Gender Reveal Material" **Context:** "Open up the package of gender reveal material that comes with the balloons. This material is usually in the form of a powder or confetti that you'll fill the balloons with. Follow the manufacturer's instructions for how much material to use and how to fill the balloons."	**Text:** "Seal the Balloons" **Context:** "Once the balloons are filled with the gender reveal material, seal them by tying off the opening with a piece of ribbon or string. Make sure the opening is securely closed so that the material doesn't spill out before it's time for the reveal."	**Text:** "Prepare the Reveal" **Context:** "Before you start the reveal, make sure you have everything you need nearby, such as a knife or pin to pop the balloons, and a container to catch the reveal material."	**Text:** "Start the Reveal" **Context:** "When it's time for the reveal, have each person hold onto their balloon and pull them apart gently. The gender reveal material will spill out of the balloons and show the gender of the baby."	**Text:** "Optional Step: Clean Up" **Context:** "After the reveal, clean up any mess that was made by disposing of the gender reveal material and inflated balloons."

Sequence 2:

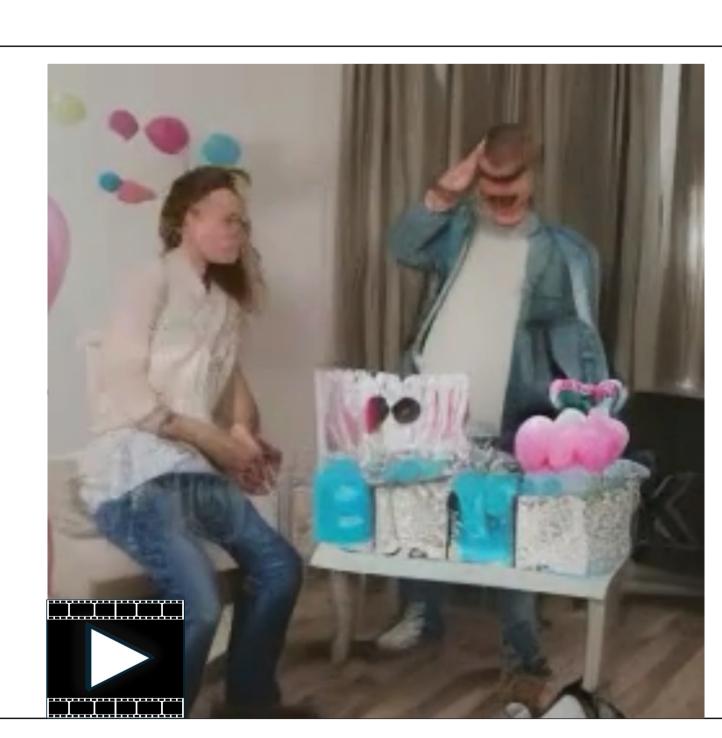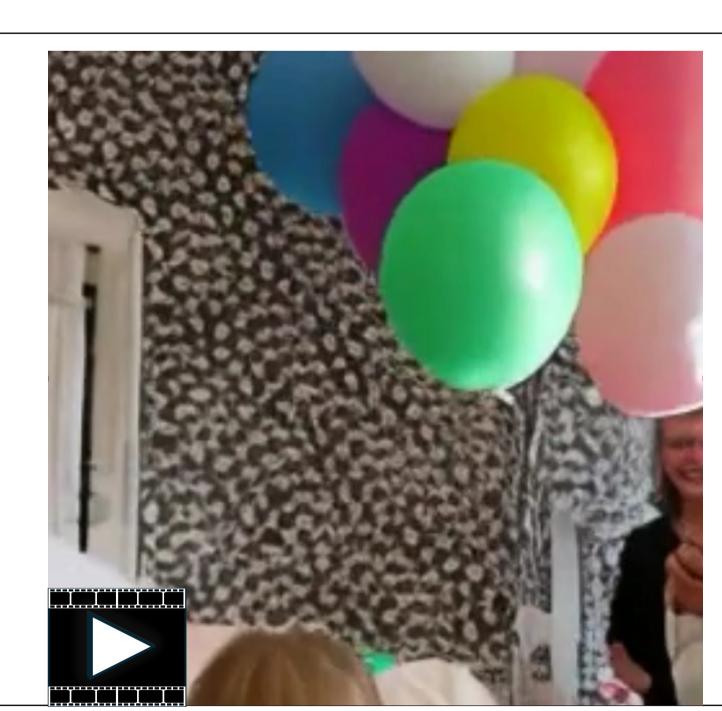| Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Text: "Materials needed" **Context:** "Gender reveal balloons (you can buy these online or at party supply stores), balloon pump, ribbon or string, tape (optional)." **Visual:** "The expectant couple is preparing the gender reveal balloon by gathering supplies and decorations."	**Text:** "Inflate the Balloons:" **Context:** "Inflate the gender reveal balloons with air using a balloon pump. Make sure to inflate them to the desired size, but not too tightly, as you want them to be easy to pop when needed." **Visual:** "They are inflating the balloon with either pink or blue confetti, depending on the baby's gender."	**Text:** "Attach the Ribbon or String:" **Context:** "Attach a ribbon or string to each balloon. This will make it easier to hold onto the balloons and pull them apart when it is time for the reveal." **Visual:** "The excited parents-to-be is sealing the balloon securely to keep the confetti hidden until the reveal moment."	**Text:** "Prepare the Reveal:" **Context:** "Before you start the reveal, make sure you have everything you need nearby, such as a knife or pin to pop the balloons, and a container to catch the reveal material." **Visual:** "They are arranging the venue and setting up the space for the gender reveal celebration."	**Text:** "Start the Reveal:" **Context:** "When it's time for the reveal, have each person hold onto their balloon and pull them apart gently. The gender reveal material will spill out of the balloons and show the gender of the baby." **Visual:** "The guests are gathering around eagerly, waiting for the big reveal with anticipation."	**Text:** "Optional Step: Clean Up" **Context:** "After the reveal, clean up any mess that was made by disposing of the gender reveal material and inflated balloons." **Visual:** "The couple is releasing the balloon into the air or popping it to reveal the pink or blue confetti."	**Text:** "Celebrate!" **Context:** "Everyone is cheering and celebrating as the gender of the baby is unveiled." **Visual:** "Everyone is cheering and celebrating as the gender of the baby is unveiled."
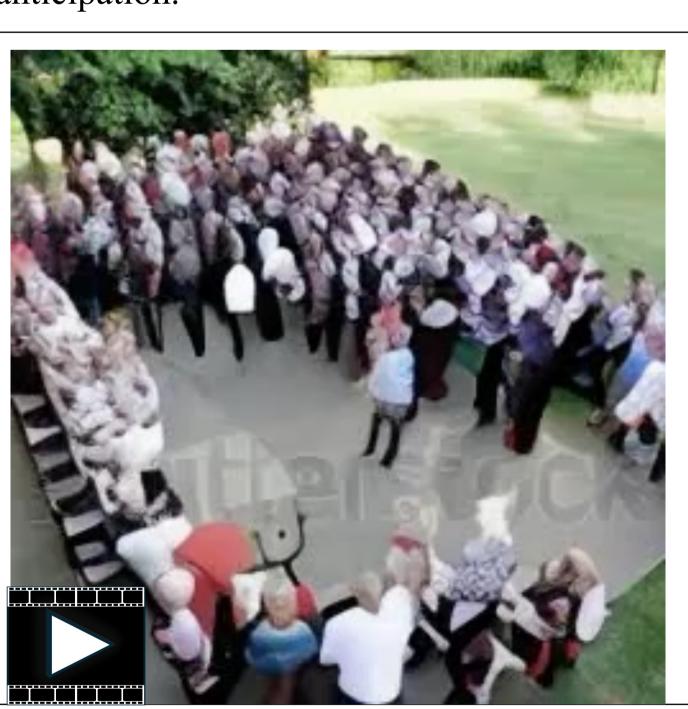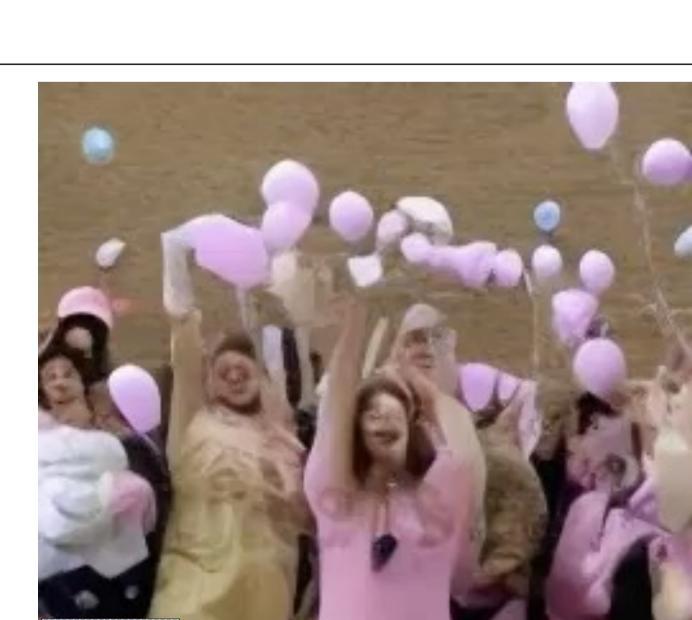
1. Which sequence contains richer <u>Textual Informativeness</u>? The richer Textual Sequence means more information in textual sequence (Step 1, Step 2, … , Step N) is for completing the given <mark>\<Task\></mark>. **Select an Option [Please only consider texts in this option.]**

○ 1 - Sequence 1 is better   ○ 0 - Tie   ○ 2 - Sequence 2 is better

2. Which sequence contains richer <u>Visual Informativeness</u>? The richer Visual Sequence means more information in video sequence (the videos under each Step) is for completing the given <mark>\<Task\></mark>. **Select an Option [Please only consider videos in this option.]**

○ 1 - Sequence 1 is better   ○ 0 - Tie   ○ 2 - Sequence 2 is better

3. Which sequence has more sensible <u>Temporal Coherence</u>? : The sensible Temporal Multimodal (text and videos) sequence follows real-world common-sense understanding. In short, the first step of <mark>\<Task\></mark> occurs before the second step instead of after. **Select an Option [Please consider texts and videos together in this option.]**

○ 1 - Sequence 1 is better   ○ 0 - Tie   ○ 2 - Sequence 2 is better

4. Which sequence is more likely to attain a greater level of <u>Plan Accuracy</u>: The greater level of Plan Accuracy means whether accomplishing the <mark>\<Task\></mark> or not. **Select an Option [Please consider texts and videos together in this option.]**

○ 1 - Sequence 1 is better   ○ 0 - Tie   ○ 2 - Sequence 2 is better