Final Report on Touch Graphics Year 2 sub-contract to DIAGRAM Center

by Val Morash, Project Evaluator

31 August 2014

The following describes a study that compares descriptions generated through templates to those generated in the traditional method of free, unstructured description based on guidelines. The majority of participants were unfamiliar with image description prior to their participation.

**Methods**

**Participants**

Twenty-two participants completed the study. Their average age was 34.81 years (SD = 13.06 years) and 6 were male.

The study began with the participants answering questions about their characteristics. The majority of participants (59%) reported having no experience describing images (for someone with a visual impairment) before this study. The other participants reported having described images in a casual context (14%), at a non-expert level for books (18%), or at an expert-level for books (9%). Most participants were non-teachers (45%), and a smaller number had taught informally as a volunteer or in the community (14%), as a classroom teacher (27%), as a special-education or paraprofessional teacher (9%), or as a teacher of students with visual impairments (5%).

An almost equal number of participants had a background in science, technology, engineering, and mathematics (STEM) as non-educators (36%), as those who had no background in STEM (32%). A fair number of participants were interested and knowledgeable about STEM with no formal background (23%), while the others taught STEM in a k-12 (5%) or post-secondary (9%) setting.

Twenty of the 22 participants were native English speakers (90%), and the other two were living and working in the United States, and reported being competent (5%) or fluent (5%) in English. Lastly, the majority of participants indicated that they were graphically literate, easily grasping concepts presented in graphics (63%), while a smaller number reported occasionally having difficulty with complex images (27%).

**Images**

 Six images of different types were selected from the National Center for Accessible Media website on guidelines for describing STEM images. These included the examples of horizontal and vertical bar charts, a line graph, a Venn diagram, a scatter plot, and a pie chart.

**Procedure**

 Participants provided responses through a free online survey tool called Lime Survey. An email was sent to each participant with a list of surveys for them to complete. The email instructed the participants that, while it was okay to take breaks between these surveys, once a survey had been started the participant should not take a break until that survey had been completed and closed.

The first survey collected demographic information from the participant (described above). The last survey was an exit survey (described below). The intervening surveys, 2-7, were the image-description surveys. The participants were informed that they were to create descriptions to make images accessible to an individual who is blind.

Each participant described all 6 images in random order, 3 randomly selected to be described without guidance and 3 with guidance/template.

 **Unguided.** During unguided description, the participant was presented with the image to be described and the NCAM guidelines for describing that type of image. The participant was given an empty text box and an empty table (sized 7 columns by 17 rows).

 **Guided/Template.** Guided descriptions presented the image, and then asked a series of questions about the image. A description and table were then auto-generated from participant answers using templates to create descriptions in-line with the NCAM guidelines. The participant was given an opportunity to edit the final text description and table.

**Exit survey.** After completing all image descriptions, the participants were asked whether they preferred description unguided, guided, or with no preference. They were also asked to rate the difficulty of describing images with and without guides on a 1-4 scale, 1 being easy and 4 being difficult.

**Analysis**

**Descriptions.** Participants’ description and table data were exported into a spreadsheet program from LimeSurvey. These were loaded into Matlab, and then output into a text file containing only the participants’ text description and table. Each description (text and table) was then evaluated along several dimensions. Average metrics for unguided and guided descriptions were compared using paired t-tests, a pair for each image type (all statistical analyses were done in R).

 The dimensions of evaluation included how long it took to complete an image description, whether the resulting description contained any text description, and the word count (evaluated in Microsoft Word) of the text description. Also, whether or not the description had a fully completed table or not. Non-completed tables resulted from guided descriptions when the participant did not answer all of the question prompts.

 The text descriptions were also evaluated for presence of syntactic (spelling, punctuation, or capitalization) errors and content (unit, number, or label) errors. Additionally, whether or not they specified the chart type and chart title, caption, units, and data/trend summary when applicable. The table descriptions were similarly evaluated for presence of syntactic and content errors.

 **Exit survey preference.** Preference scores from each participant, ranging from 1=easy to 4=difficult, for unguided and guided descriptions were compared using a paired t-test. The participants had also indicated whether they preferred writing image descriptions unguided, guided, or had no preference. These ratings were evaluated by a Thurstonian model, which estimated the underlying distributions of preference for unguided and guided description and then tested whether the difference in means between these distributions, scaled by standard deviation, commonly referred to as d’, was significantly different from zero (Christensen, Lee, & Brockhoff, 2012). If significant, there is statistical evidence that one of the types of description was preferred over the other. This analysis was done in R using the sensR package (Christensen, et al., 2012).

**Results**

Standard errors in parentheses.

**Overall**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Unguided | Guided | Test Statistic | *p* |
|  |  |  |  |  |
| **Text and Table** |  |  |  |  |
| Time to Describe (sec) | 738.26 (155.21) | 613.19 (111.05) | t(5) = 1.20 | 0.282 |
| Any Text Description (%) | 96.97 (3.03) | 100.00 (0.00) | t(5) = -1.00 | 0.363 |
| Text Word Count (count) | 63.09 (9.28) | 82.17 (20.23) | t(5) = -1.19 | 0.289 |
| Completed Table (%) | 60.00 (10.98) | 89.09 (3.40) | t(4) = -2.14 | 0.099 |
|  |  |  |  |  |
| **Text Description** |  |  |  |  |
| Syntactic Error (%) | 37.04 (7.49) | 48.48 (13.42) | t(5) = -0.78 | 0.469 |
| Content Error (%) | 11.28 (4.62) | 12.12 (5.59) | t(5) = -0.219 | 0.835 |
| Trend/Data Summary (%) | 61.28 (11.42) | 75.76 (16.87) | t(5) = -1.05 | 0.343 |
| Chart Type (%) | 72.05 (6.49) | 100.00 (0.00) | t(5) = -4.31 | 0.008 \*\* |
| Title (%) | 34.55 (15.05) | 83.64 (16.36) | t(4) = -2.90 | 0.044 \* |
| Caption (%) | 18.18 (5.25) | 93.94 (3.03) | t(2) = -9.45 | 0.011 \* |
| Describe Units (%) | 36.20 (6.59) | 100.00 (0.00) | t(5) = -9.67 | < 0.001 \*\*\* |
|  |  |  |  |  |
| **Table** |  |  |  |  |
| Syntactic/Style Error (%) | 12.66 (3.58) | 8.44 (3.81) | t(4) = 1.64 | 0.177 |
| Content Error (%) | 14.52 (6.65) | 10.22 (5.48) | t(4) = 0.644 | 0.554 |
|  |  |  |  |  |
| **Participant Feedback** |  |  |  |  |
| Difficulty (1-4) | 2.95 (0.18) | 1.95 (0.14) | t(21) = 5.07 | < 0.001 \*\*\* |
| Preference (count) | 3 | 16 (3 no pref.) | d' = 1.20 (0.40) | 0.002 \*\* |
|  |  |  |  |  |
| Paired t-tests, \* p < 0.05, \* p < 0.01, \* p < 0.001 |  |  |  |  |

**Discussion**

 Guided and unguided descriptions were equally likely to contain syntactic or content errors, contain text or table elements, and provide a description of the data and data trends. Guided and unguided descriptions were also equal in how long they took to create and their text word counts. However, guided descriptions were more likely to contain important image information, including the chart type, title, caption, and units. Furthermore, the participants found description using guidance/templates to be significantly easier, and they preferred this method over unguided description.

**Recommended Future Directions**

 The current report assesses the production of image descriptions through unguided and guided interfaces, and compares the resulting descriptions on their accuracy and completeness. The guided descriptions were no more or less accurate than the unguided descriptions, but were more complete. The current research does *not* address whether unguided or guided descriptions promoted comprehension and usability. Further research will be needed to investigate what properties of descriptions work best for students and teachers, and whether these properties are encouraged or discouraged by our guided-description system.

**References**

Christensen, R.H.B., Lee, H-S & Brockhoff, P.B. (2012). Estimation of the Thurstonian model for the 2-AC protocol. Food Quality and Preference, 24, 119–128.