Users' Interpretation of *Pictograms* and *Pictures* for Conveying Instructions and Warnings on Pharmaceutical Labels

Destyn Jones^{1(\boxtimes)}, Sabrina Moran^{2(\boxtimes)}, Jaime Sanchez^{2(\boxtimes)}, Amber Latham^{2(\boxtimes)}, and Kim-Phuong L. Vu^{1,2(\boxtimes)}

¹ Building Infrastructure Leading to Diversity (BUILD) Program, California State University, Long Beach, Long Beach, CA 90815, USA Destyn. Jones@student.csulb.edu, Kim.Vu@csulb.edu ² Center for Usability in Design and Accessibility (CUDA), California State University, Long Beach, Long Beach, CA 90815, USA Sabrina.N.Moran@gmail.com, AmberLatham@gmail.com, Jaime7529@att.net

Abstract. Pharmaceutical pictograms are figures that provide a visual representation of medication instructions and warnings. Pictograms were created in order to lower the continuously rising problem of poor medication adherence due to patients misunderstanding what is being instructed on medication labels. Research on pictograms show that their use could increase patients' comprehension of medication instructions, but the comprehension rates are still low. Researchers have argued that in order to improve the effectiveness of pictograms, they must be more concrete, simple, meaningful, and familiar. Pictures have these properties. Thus, the goal of this study was to examine if the use of pictures (i.e., real images) provide a better representation of what is being instructed to users on medication labels than pictograms. We hypothesized that real images would lead to a better understanding of the instructions and warnings for taking medication, and in turn, users would have higher comprehension levels. Participants completed an online survey evaluating the meaning of either pictograms or pictures. Both quantitative and qualitative data analyses were conducted on free responses to the questions and on users' ratings of the "goodness" of the stimuli. The results showed that participants were moderately accurate (74%) in their interpretation of pictograms, and the use of pictures did not improve their comprehension.

Keywords: Pictograms · Pictures · Real images · Medication instructions · Medication adherence · Comprehension levels

1 Introduction

According to the Institute of Medicine, nearly every individual living today has taken or will take medication at some point in their lifetime. Specifically, approximately four out of five U.S. adults will use prescription medicines, or other drugs received over-thecounter, on a weekly basis [1]. These medications are meant to be beneficial to consumers; however, they can be harmful if the medication is not taken appropriately [2].

© Springer Nature Switzerland AG 2019

S. Yamamoto and H. Mori (Eds.): HCII 2019, LNCS 11569, pp. 22–33, 2019. https://doi.org/10.1007/978-3-030-22660-2_2 Errors in taking prescription medicine are more common than expected. For example, Brown and Bussel [3] indicated that 50% of people who are prescribed pharmaceutical prescriptions do not take them correctly. Of these people, 58% of these errors are due to inadequate medication labeling [4]. In efforts to prevent errors caused by insufficient labeling, the Food and Drug Administration (FDA) established guide-lines for the labeling of substances and filling prescriptions. Under Title XXI, companies are required to include the pharmacy name, address, prescription serial number, date of initial filling, name of the patient, name of the practitioner issuing the prescription, drug name and dosage, as well as the directions for use and cautionary statements on all medication labels [5]. This information is helpful; however, it may not be enough to allow users to understand all of the information they need to know regarding the medication. In order to further build upon the standards set by the FDA, Pharmacopeia and the Institute for Safe Medicine Practices suggested that labels should contain:

- 1. Words that are typed and easy-to-read in 12-point font
- 2. Warnings that are typed directly onto labels in large font
- 3. The generic and brand name of the drug
- 4. Images or physical descriptions of the pills in the container
- 5. No extra zeros
- 6. Pharmacy information at the bottom of the label

Herron and Vu [6] conducted a study to determine whether the Pharmacopeia and the Institute for Safe Medicine Practices recommendation for adding images or physical descriptions of the pills in the container would help people in identifying the correct medication to take. Herron and Vu included conditions wherein participants were asked to identify the medication in a container by viewing labels that had a black and white picture of the medicine, a color picture of the medicine, or text description of the medicine. They found that participants were faster and more accurate at identifying the correct medication with the labels that had a color depiction of the medication compared to ones that had a text description or a black and white picture of the medicine. Based on this finding, they recommended that if a pharmacy were to employ pictures of the medication, they do so only if they can provide color pictures of the medicine on the labels.

One reason that pictures and symbols are recommended on pharmaceutical labels is that illiterate patients are 3.4 times less likely to comprehend prescription labels correctly [7]. Patients' inability to understand the instructions given on pharmaceutical labels compromises the medications' outcomes and increases the risk of side effects. To improve patients' understanding of pharmaceutical instructions, pictograms were created. Pictograms give a visual representation of the instructions listed on pharmaceutical labels. Although, people interpret pictograms with higher accuracy than text [7], pictograms have not been shown to result in high levels of comprehension [8]. Nevertheless, as previously noted, pictograms can be particularly beneficial for users with low literacy.

To make pictograms more effective, pharmaceutical pictograms should be visible and easy to interpret, make direct connections to familiar things, be concrete [8], and be semantically close [9]. Semantic closeness can improve interpretations of pictograms by making them more accessible and relatable [9]. Furthermore, the picture superiority effect [10] states that pictures are more likely to enhance memory than words through top-down processing. Top-down processing theorizes how individuals use their general knowledge to understand new information. Pictures are commonly used amongst individuals; therefore, top-down processing should allow them to better understand what is being depicted in pictures. As a result, pictures are used in many contexts as memory aids.

The purpose of this study was to examine if real images that depict instructions and warnings lead to better comprehension of what is being instructed to patients than pictograms. Participants were recruited from Amazon Mechanical Turk (mTurk) to complete an online survey in which they interpret the meaning of a pictogram or picture that depicts a medical instruction or warning that would be found on a pharmaceutical label. After participants provided their interpretation of the stimuli, the intended meanings of the pictograms or pictures were provided, and participants were asked to rate the stimuli's ability to capture the intended meaning. It was hypothesized that participants would be more accurate in their interpretation of the intended meaning of the instruction or warning when the stimuli were pictures rather than pictograms.

2 Methods

2.1 Participants

Overall, 236 participants were recruited from mTurk to participate in this study. Participants were U.S. residents who were 18 years of age or older, self-identified as fluent in English, and had a 95% or higher approval rating on mTurk. The demographics for participants in each experimental condition are described below for each survey type.

Pictogram Survey. There were 120 participants who completed the pictogram survey, $(M_{age} = 36.40, SD = 11.36)$. Seventy-six were female (63%) and 44 were male (37%). Ninety-one of the participants self-identified as White (76%), 10 as Hispanic (8%), 14 as Asian (12%), and five as Black or African American (4%). Sixty participants reported that their yearly income was over \$50,000 a year (50%), with seven specifically reporting their income was over \$100,000 a year (12%). Twenty-four participants had a high school diploma or equivalent (20%), 28 had an Associate's degree (23%), 55 had a Bachelor's degree (46%), and 13 had a Master's degree (11%).

Picture Survey. There were 116 participants who completed the picture survey, $(M_{age} = 37.74, SD = 11.48)$. Sixty-four were female (55%) and 52 were male (45%). Ninety participants self-identified as White (78%), six as Hispanic (5%), seven as Asian (6%), eight as Black or African American (7%), and five as other (Native American, Pacific Islander, etc.; 4%). Fifty-two participants reported that their yearly income was over \$50,000 (45%); out of those participants, eight reported a yearly income of over \$100,000 (16%). Seventeen participants had a high school diploma or equivalent (15%), 16 had some four-year college (14%), 19 had an Associate's degree (16%), 48 had a Bachelor's degree (41%), 10 had a Master's degree (9%), and six had a Ph.D. or M.D (5%).

2.2 Design

The study utilized a between-subjects design, with survey type (Pictogram vs. Picture) as the independent variable. The dependent variables were the accuracy scores in interpreting the stimuli and rating scores for the "goodness" of the stimuli. Accuracy scores were determined in two ways. First, responses were classified as correct, partially correct or incorrect. Second, a numeric value was assigned to the answer given by the participants based on their accuracy in generating the intended meaning of the picture or pictogram. Rating scores reflected how well the picture or pictogram aligned with its intended meaning once participants were told its meaning.

2.3 Materials

Pictogram Stimuli. The pictograms used in the study were extracted from the U.S. Pharmacopeial Convention (USPC). The USPC is "a scientific nonprofit organization that sets standards for the identity, strength, quality, and purity of medicines, food ingredients, and dietary supplements manufactured, distributed and consumed worldwide" [5]. Enforced by the FDA, the USPC creates standard pharmaceutical pictograms that companies can use on their medication labels in order to convey their medication instructions. Not all companies utilize the USPC's pictograms, as they are not mandated to do so. However, if companies choose to include pictograms, they must select the pictograms from the selection offered through the USPC.

Picture Stimuli. The pictures used in the study were gathered from Google images or taken by experimenters. Picture stimuli were manipulated in order to be as close as possible to their pictogram counterpart (see example below in Fig. 1).



Fig. 1. An example of a picture and pictogram stimuli portraying "drink additional water."

Surveys. Two versions of an online survey were constructed in Qualtrics, an online survey platform. Each survey consisted of 3 blocks of questions. The first block included open-ended questions that asked, "What does the following image mean to you?" and was presented with a stimulus (either a pictogram or picture, depending on the survey type). Participants responded by typing answers into a response box. The second block consisted of rating questions. The rating questions asked participants to rate from 1 to 5, with 1 being "strongly disagree" and 5 being "strongly agree,"

how well the stimuli (i.e., pictogram or picture) aligned with its intended meaning. Each question was presented one at a time in randomized order within the block. The last block consisted of demographic questions. On average, it took participants 30 min to complete each survey. Data collection was completed in one day.

2.4 Data Coding

In order to compute the accuracy of the participants' responses to the stimuli, a content analysis was performed on the open-ended responses. Participants' responses were grouped based on similarities in semantic meaning. These groups of responses were then classified as being correct, partially correct, or incorrect. Whether a response was correct, partially correct, or incorrect was determined by an agreement across eight independent raters. For example, for the stimuli depicting: "if this medication makes you dizzy, do not drive," correct responses included, "medication may make you dizzy, do not operate motor vehicles after taking" and "do not drive if this medication makes you dizzy." Examples of partially correct responses for this stimulus included: "do not operate heavy machinery," "do not drive," "potential dizziness," or "may cause dizziness." These responses were considered partially accurate because only one aspect of the intended meaning of the stimuli was stated. As another example, for the stimuli depicting, "take with additional water," correct responses included: "drink two glasses of water with medication" and "take with plenty of water." Responses that stated, "take with water" were considered partially correct because these responses do not indicate drinking more or "plenty" of water (which the raters defined as drinking more than one glass of water).

Interrater reliability was determined by the percentage of agreement amongst all raters in deciding whether participants' responses to each stimulus was correct, partially correct, or incorrect. The percent of questions with agreed coding from all raters was 79% for the picture survey and 70% for the pictogram survey. It should be noted that this level of agreement is considered acceptable given that agreement was determined across eight independent raters. All inconsistent classifications were discussed and classified by consensus of the raters. To compute an accuracy score, correct responses received a score of 1, partially correct responses received a score of 0.5, and incorrect responses received a score of 0.

3 Results

3.1 Accuracy

Accuracy was analyzed in two ways: one based on the frequency of correct, partially correct, and incorrect classifications through a chi-square analysis, and one based on an accuracy score through t-tests.

For the frequency analysis, a chi-square test was performed on the 26 stimuli that shared the same meaning between the picture and pictogram surveys to see if the frequency of response accuracy classification differed between the two surveys. Results showed that there were significant differences in response categorization for 18 stimuli,

| Stimulus # and | Pictogram | Picture | Test Statistics |
|----------------------------------|------------------------------|---------------------------|--|
| Meaning #2 Avoid Salt | | 2 | |
| | | | |
| | | S | |
| | C = 80.49% | C = 18.33% | $X^2(2, N=243) =$ |
| | PC = 12.19% | PC = 73.33% | 100.77, <i>p</i> < .001 |
| | $M_{\rm score} = 0.53$ | $M_{\rm score} = 0.55$ | <i>t</i> (241) = 0.70, p = .484 |
| #3 Do not crush, | | - | |
| break, or open capsule | | | |
| | | 6 | |
| | C = 83.74% | C = 69.17% | $X^{2}(2, N = 243) =$ |
| | PC = 5.69% | PC = 14.17% | 7.77, <i>p</i> = .021 |
| | $M_{\rm score} = 0.87$ | $M_{\rm score} = 0.76$ | <i>t</i> (241) = -2.28, p = .023 |
| #4 This medicine may make you | A CLE | | |
| dizzy, do not drive | $\langle \mathbf{x} \rangle$ | | |
| | C = 60.16% PC = 37.40% | C = 19.17% PC = 73.33% | $X^{2}(2, N = 243) =$ 42.95, $p < .001$ |
| | $M_{\rm score} = 0.79$ | $M_{\rm score} = 0.56$ | <i>t</i> (241) = -6.84, p < .001 |
| #5 Do not share with others | | | |
| with others | | | |
| | C = 39.84% | C = 40.00% | X^2 (2, $N = 243$) = |
| | PC = 30.89% | PC = 2.50% | 40.23, <i>p</i> < .001 |
| | $M_{\rm score} = 0.55$ | $M_{\rm score} = 0.41$ | <i>t</i> (241) = -2.42, p < .016 |
| #7 Take in the mornings | | 30 | |
| monnings | | | |
| | C = 68.30% PC = 25.20% | C = 84.17% PC = 0.00% | $X^{2}(2, N = 243) =$ 37.01, $p < .001$ |
| | $M_{\text{score}} = 0.81$ | $M_{\text{score}} = 0.84$ | t(241) = 0.76, p = .449 |

Table 1. The significant difference between pictures and pictograms sharing the same meaning. C = correct; PC = partially correct; $M_{\text{score}} = \text{mean accuracy score}$

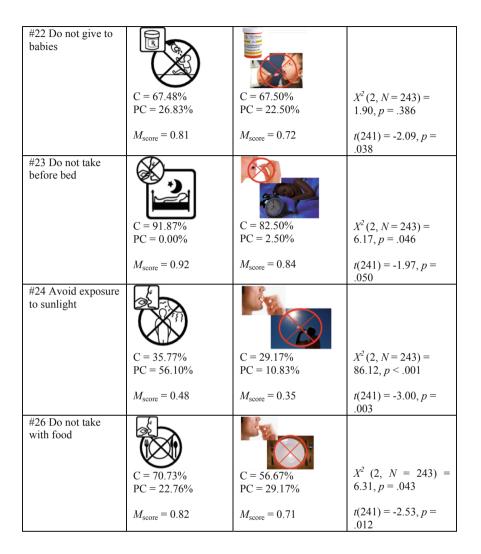
(continued)

| #9 Take by mouth | | | · · · · · · · · · · · · · · · · · · · |
|--|--|---|--|
| | C = 65.04% PC = 24.39% | C = 45.00% PC = 34.17% | X^2 (2, N = 243) = 10.50, p = .005 |
| | $M_{\rm score} = 0.77$ | $M_{\rm score} = 0.62$ | <i>t</i> (241) = -3.24, p = .001 |
| #10 Check pulse | | | |
| | C = 85.36% PC = 1.63% | C = 70.00% PC = 10.00% | X^2 (2, $N = 243$) = 11.04, $p = .004$ |
| | $M_{\rm score} = 0.86$ | $M_{\rm score} = 0.75$ | <i>t</i> (241) = -2.33, p = .021 |
| #13 Store medica- tion in the refrigera- tor | C = 91.06% PC = 0.81% M _{score} = 0.91 | C = 74.17% PC = 0.00% $M_{score} = 0.74$ | $X^{2}(2, N = 243) =$ 14.53, p = .001 t(241) = -3.68, p < .001 |
| #15 Take three times a day with food | C = 69.92% PC = 24.39% $M_{\text{score}} = 0.82$ | $C = 70.83\%$ $PC = 15.00\%$ $M_{score} = 0.78$ | $X^{2}(2, N = 243) =$ 7.14, p = .028 t(241) = -2.28, p <.023 |
| #16 Take two hours after eating | C = 92.68% PC = 2.44% | C = 85.45% PC = 0.91% | $X^2(2, N = 243) =$ 6.07, p = .048 |
| | $M_{\rm score} = 0.94$ | $M_{\rm score} = 0.79$ | <i>t</i> (241) = -0.89, p = .375 |

(continued)

| #17 Take an hour before eating | | | |
|--|---------------------------|---------------------------|---|
| | C = 80.49% PC = 7.32% | C = 67.50% PC = 2.50% | $X^{2}(2, N = 243) =$ 13.41, p = .001 t(241) = -2.98, p = |
| | $M_{\rm score} = 0.84$ | $M_{\rm score} = 0.69$ | .003 |
| #18 Do not give to children | | | |
| | C = 75.61% PC = 17.89% | C = 60.83% PC = 21.67% | $X^{2}(2, N = 243) =$ 8.54, $p = .014$ |
| | $M_{\rm score} = 0.85$ | $M_{\rm score} = 0.78$ | <i>t</i> (241) = -1.62, p = .106 |
| #19 Do not refrig- erate | | | |
| | C = 85.36% PC = 4.88% | C = 67.50% PC = 3.33% | $X^{2}(2, N = 243) =$ 14.72, $p = .001$ |
| | $M_{\rm score} = 0.88$ | $M_{\rm score} = 0.69$ | <i>t</i> (241) = -3.75, <i>p</i> < .001 |
| #20 Dissolve in water | C = 82.93% PC = 0.00% | C = 66.67% PC = 0.00% | $X^{2}(2, N = 243) =$ 8.54, p = .003 |
| | $M_{\rm score} = 0.83$ | $M_{\rm score} = 0.67$ | t(241) = -2.96, p = .003 |
| #21 This medica- tion may cause drowsiness | | | |
| | C = 85.36% PC = 9.76% | C = 80.83% PC = 0.83% | $X^{2}(2, N = 243) =$ 18.73, $p < .001$ |
| | $M_{\rm score} = 0.88$ | $M_{\rm score} = 0.81$ | <i>t</i> (241) = -1.46, p = .147 |

(continued)



which are shown in Table 1. For the pictogram survey, the percent of correct responses was 74.33%, with 16.34% classified as being partially correct and 9.33% as incorrect. For the picture survey, the average correct response was 60.81%, with 16.67% classified as partially correct and 22.52% classified as incorrect. Thus, participants were more accurate with interpreting the stimuli in the pictogram survey than the picture survey.

Accuracy scores for the pictogram and picture stimuli were further compared using independent-sample t-tests (see last column in Table 1 for accuracy scores). The overall accuracy score was 0.80 for the pictogram survey and 0.69 for the picture survey. The accuracy score for 14 stimuli were significantly different based on the survey type.

31

Pictograms had higher accuracy scores than pictures for 12 of the 14 stimuli (86%); pictures had a higher accuracy than pictograms for 2 of the 14 stimuli (14%).

The results from the chi-square and t-test analyses converged. The chi-square tests were more sensitive and found more differences in stimuli responses between the survey types (pictogram vs. picture) than the *t*-tests revealed.

3.2 Rating Scores

One-way analyses of variance (ANOVAs) were performed for each stimulus to determine participants' ratings of how well the pictures or pictograms aligned with their intended meaning, see Table 2. The independent variable was the stimuli type (pictogram or picture) and the dependent variable was the rating score that each stimulus received. Overall, participants rated the stimuli as being good, with the average rating for pictograms being 4.57 out of 5, and the average rating for pictures being 4.54 out of 5. Nine stimuli showed significant differences in participants' rating scores as a function of the survey type. The pictogram was rated to be better than the picture in 8 out of 9 cases (89%).

| Stimulus # and meaning | Mean rating | F-statistic |
|--|--|-----------------------------|
| #5 Do not share with others | Pictogram $M = 4.30$ Picture $M = 3.38$ | F(1, 237) = 25.61, p < .001 |
| #13 Store medication in the refrigerator | Pictogram $M = 4.75$ Picture $M = 4.43$ | F(1, 238) = 7.49, p = .006 |
| #17 Take an hour before eating | Pictogram $M = 4.67$ Picture $M = 4.37$ | F(1, 239) = 6.29, p = .013 |
| #18 Do not give to children | Pictogram $M = 4.66$ Picture $M = 4.43$ | F(1, 237) = 3.88, p = .050 |
| #20 Dissolve in water | Pictogram $M = 4.63$ Picture $M = 4.29$ | F(1, 238) = 6.77, p = .010 |
| #22 Do not give to babies | Pictogram $M = 4.61$ Picture $M = 4.19$ | F(1, 239) = 11.59, p = .001 |
| #23 Do not take before bed | Pictogram $M = 4.63$ Picture $M = 4.27$ | F(1, 238) = 7.33, p = .007 |
| #24 Avoid exposure to sunlight | Pictogram $M = 4.22$ Picture $M = 3.33$ | F(1, 238) = 30.20, p < .001 |
| #26 Do not take food | Pictogram $M = 4.66$ Picture $M = 4.38$ | F(1, 237) = 4.83, p = .029 |

Table 2. The rating scores for the effectiveness of the picture and pictogram stimuli performed through an ANOVA test.

4 Discussion

The purpose of this study was to determine if real images would provide users with a better understanding of medication instructions and warnings than pictograms. Contrary to the hypothesis, participants were better at interpreting the meaning of pictograms compared to pictures. However, participants' comprehension for pictogram stimuli were still low, with only 74% of responses being classified as correct interpretations.

A potential reason why pictures were not more understandable than the pictograms could be that when viewing the pictures, participants fixated on the specific details presented in the pictures which led them to more specific, literal interpretations. This was detrimental as some of the pictures used in the study contained additional, irrelevant details that may have served as distractors. For example, for the picture stimuli, "avoid exposure to sunlight," a red "x" covers the picture of a man dressed in active wear, drinking water in the sun. Due to the picture depicting a man wearing active wear outdoors, many participants focused on this irrelevant cue when interpreting the meaning of the picture. This was seen in participants' responses, with some being, "do not drink fluids or be outside when taking medication", "do not exercise after taking this," and "do not take with water." Similar misinterpretations occurred with the picture stimuli, "do not share with others," as the glare from the sun made the medication bottle and pills used in the stimuli hard to recognize. As a result, some of the responses received were "light prohibited", "do not apply it to your skin," "do not play with fire," and "do not touch hot liquid." Thus, due to the lack of clarity in the picture stimuli selection, the benefits of the pictures were reduced. Future research should focus on developing better, more simplistic picture stimuli that do not have any distractor. In addition, future research could investigate whether individuals would understand pictures better than pictograms if the pictures were generated to depict the intended meaning rather than be replications of pictograms. These improvements to the picture stimuli could potentially render higher comprehension levels for pictures.

Acknowledgements. Gratitude is extended to graduate students in the Center for Usability in Design and Accessibility (CUDA) lab for assisting with the qualitative data coding. Also, the National Institute of Health (NIH) (grant numbers: 5UL1GM118979; 5TL4GM118980; 5RL5GM118978), for providing the funding needed to carry out this research.

References

- Health and Medicine Division (2015). http://www.nationalacademies.org/hmd/Reports/ 2006/Preventing-Medication-Errors-Quality-Chasm-Series.aspx. Accessed 10 Apr 2018
- McGinley, P., Rashidee, A., Hart, J., Chen, J., Kumar, S.: Data trends: high-alert medications: error prevalence and severity (2009). https://www.psqh.com/analysis/datatrends-july-august-2009/. 10 Apr 2018
- Brown, M.T., Bussell, J.K.: Medication adherence: WHO cares? (2011). https://www.ncbi. nlm.nih.gov/pmc/articles/PMC3068890/. Accessed 23 Oct 23 2017

- Jeetu, G., Girish, T.: Prescription drug labeling medication errors: a big deal for pharmacists (2010). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035877/. Accessed 27 Sept 2018
- Work at the United States Pharmacopeial Convention. (n.d.). https://www.careerbuilder. com/company/the-united-states-pharmacopeial-convention/C8S05668H22YDHCWJ0Y. Accessed 10 April 2018
- Herron, M., Vu, K.L.: The value of including a picture of the medicine on pharmaceutical labels. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 57(1), 688–692 (2013). https://doi.org/ 10.1177/1541931213571149
- Davis, T.C.: Low literacy impairs comprehension of prescription drug warning labels (2006). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1831578/. Accessed 10 Apr 2018
- Chan, A.H., Chan, K.W.: Effects of prospective-user factors and sign design features on guessability of pharmaceutical pictograms. Patient Educ. Counsel. 90(2), 268–275 (2013). https://doi.org/10.1016/j.pec.2012.10.009
- Dowse, R., Ehlers, M.: Pictograms for conveying medicine instructions: comprehension in various South African language groups (2004). https://journals.co.za/content/sajsci/100/11-12/EJC96165. Accessed 17 Oct 2017
- Stenberg, G.: Conceptual and perceptual factors in the picture superiority effect. Eur. J. Cognit. Psychol. 18(6), 813–847 (2006). https://doi.org/10.1080/09541440500412361