Improving Individual and Team Decisions Using Iconic Abstractions of Subjective Knowledge

Robert A. Fleming
SPAWAR Systems Center
Code 24402
53560 Hull Street
San Diego, CA 92152
619.553.3628
bob.fleming@navy.mil

Michael B. Cowen
SPAWAR Systems Center
Code 244210
53560 Hull Street
San Diego, CA 92152
619.553.8004
mike.cowen@navy.mil

Improving Individual and Team Decisions Using Iconic Abstractions of Subjective Knowledge

Robert A. Fleming
Michael B. Cowen
SPAWAR Systems Center, San Diego

Abstract

Eighteen military officers made ratings about the suitable location of a mythical refugee camp. They were given 30 information items, six items for each of five decision criteria. Using a 5-point acceptability scale, they gave a rating for each of the criteria plus an overall acceptability rating. Half of the participants (Text group) saw each item and assigned it to a criterion, and then during the decision phase were shown the items they had assigned to each criterion and asked to make their ratings. The other half of the participants (IOB group) had the additional task of evaluating each item on a 4-point yes-no rating scale. During the decision phase they were not shown the original text items but rather an information object (IOB) representation of the item, (i.e., a small symbol which encapsulated their subjective rating of the item). Half of the participants should have made a Positive recommendation, half a Negative recommendation. The results showed that not a single participant in the Text group made a correct overall assessment, predominately selecting the "neutral' option. Six of the nine IOB participants made the correct overall assessment. Both groups took significantly longer to process information and make ratings when the preponderance of the information was positive in content versus negative. Results are discussed in terms of the increased confidence the IOB participants had in the accuracy of their ratings and the increased time needed to review and assess positive information that endorse commitment to an action.

Introduction

The work of Stasser and his associates (Stasser & Titus, 1985; Wittenbaum & Stasser, 1996; Wittenbaum, 1998) have clearly shown two significant problems in team decision making. The first is that participants are very poor at sharing their uniquely held information, even though it may have a significant influence on the final decision. They do not intentionally hide the information, group discussions center on the commonly held shared information, and due to time constraints, the uniquely held information is not shared. Even if one can devise a way to share this uniquely held information, it is often discounted by the group and will not impact the final decision. This later phenomenon is discussed by Dennis (1996) in an article aptly entitled "You Can Lead a Group to Information but You Can't Make It Think." A more complete review of this literature has been recently reported by Fleming & Kaiwi (2002).

We hypothesize that uniquely held information is not shared because it increases the cognitive burden of the recipient. In a distributed group decision making environment, each team member must collect, analyze and integrate his or her information. Processing information sent by another is an additional cognitive burden. It is likely to be ignored simply because the recipient does not have the time to analyze and integrate the information. The information has, however, already been analyzed and evaluated by the sender. The sender was motivated to share the information after assessing the information quality, importance, and effect/impact. These subjective assessments are not typically sent with the raw document.

We propose an approach where information (e.g., text documents, message traffic, etc.) is converted into iconic representations that encapsulate an individual's subjective perception of the information. These icons are referred to as Information Objects (IOBs) and are automatically generated from an abstraction form completed by a team member for a relevant information item. The small icon is analogous to a tactical symbol, but represents a document rather than sensor data. We feel that it is far more likely to be used by the recipient than a raw, unprocessed text document. A more detailed discussion of the concept of IOBs is presented in Fleming (2003).

This paper reports an experiment that evaluated the use of IOBs in an individual (vice group) decision making environment. It is logical to first demonstrate that IOBs can improve individual decision making as a prerequisite for improved group decision making. Issues specifically addressed were: (1) Does the abstraction process create an undue cognitive burden on the individual? (2) Do IOBs improve decision quality above and beyond the decisions made with original text information? (3) How do users feel about using IOBs in their decision process?

Method

Participants

The participants were 18 military officers from the Naval Postgraduate School in Monterey, CA. They were asked to participate as part of the classroom requirement. The participants were tested individually on the internet. Each participant was sent a password by email, and participated in the experiment at their convenience.

Scenario/Task

Once the participants logged onto the web site and entered their password, an instructional page appeared and tasking them to decide if the mythical country of Islandia was a good site for a refugee camp needed to house people fleeing from a volcanic eruption. They were to base their decision on five criteria: Communications Facilities, Transportation Facilities, Labor Pool, Administrative Requirements and Sanitation/ Health/Medical conditions. The tasking informed the participants that they would be presented 30 information items about Islandia and they were to assign each to the most appropriate criterion. After assigning the 30 items, they would score each criterion on a 5-point scale ranging from "very negative" to "very positive". Finally, they made an overall rating using the same scale.

Basic Design

The experimental design was a 2 X 2 between subjects factorial and is shown in Table 1. The two independent variables were Decision and Display. The preponderance of the information for **deciding** if Islandia should house a refugee site was either **Positive** (supporting Islandia) or **Negative** (not favoring Islandia). The 30 sorted information items were **displayed** either as raw **text only** or as **IOBs**.

The Decision Variable:

There was a total pool of 40 information items, eight items for each of the five criteria. For each criterion four of the eight items were supportive (positive) of Islandia, while the other four were

critical (negative). For each participant the 30 items were randomly selected from this pool such that six information items were presented for each of the five criteria. Each criterion always had a 4-2 split in terms of positive and negative information items. If a criterion had four positive and two negative, it was considered a Positive criterion, if it had four negative and two positive it was considered a Negative criterion.

Positive: In the Positive experimental condition, three of the five criteria were positive and should result in an overall positive (supportive) decision about using Islandia as the refugee site.

Negative: In the Negative experimental condition, three of the five criteria were negative and should result in an overall negative (non-supportive) decision about using Islandia.

There were ten participants in the Positive condition and eight participants in the Negative condition.

The Display Variable:

Text Group: After being presented the initial instructional page about tasking, the Text participants were presented a sample information item and instructions on how to assign it to a specific criterion. An example of one of our information items is shown in Figure 1. Note that the criterion definitions were always available for this task.

After the participant assigned the 30 items to the most appropriate criterion, the participant was presented all the items, sorted by criterion, and instructed on how to score each of the criterion as well as the overall assessment. An example of a scoring page is shown in Figure 2.

IOB Group: After being presented the initial instructional page about tasking, the IOB participants were presented a sample information item and how to assign it to a specific criterion. This is shown in Figure 3. Additional information was requested about the effect of the item on the criterion (which was needed to create IOBs). For each information item they were required to enter a keyword, assign the item to a criterion (same task as the Text subjects) and finally, assign on a 4-point scale, how supportive is the item with respect to the assigned criterion. They were provided a completed example, which is shown in Figure 4. Here, the participant read the item, used "ports" as a keyword, assigned it to the transportation criterion (TSP) and felt that it had a "positive" supporting effect upon that criterion.

After the IOB participant had completed the 30 items, the participant was presented information describing the concept of IOBs, how their inputs had been incorporated into the IOB, and how they could sort and examine the IOBs. The participant then scored each criterion and the overall assessment. An example of the scoring page is shown in Figure 5. After each of the IOB participants scored the five criteria and the overall assessment, each was asked to assess the value of having IOBs available during their decision phase by rating the IOBs on a 4-point scale (i.e., "Somewhat Distracting", "No Effect", "Helped Somewhat" and "Helped a lot").

There were nine participants in the Text group, and nine participants in the IOB group.

Results

Item Validation

In a previous pilot study, our pool of 40 information items was validated by 12 associates. Any items, which were assigned to wrong criteria or to a wrong effect by more than two members, were either dropped or modified.

Our participants made some assignment errors: Text and IOB groups had 17 and 14 errors respectively. Since there were a total of 270 possibilities for differences (9 X 30), this is about a 5%-6% rate of assigning the items to the wrong criterion. The IOB participants incorrectly rated 8 items in terms of effect/impact (positive versus negative impact), about a 3% level of disagreement.

Use of Scoring Options: We were interested in the distribution of ratings assigned to the criteria and the overall assessment. The results (see Figure 6) for all the criteria show a typical "bell" distribution, with no real differences between the Text and IOB conditions. It is interesting to note the high frequency of the neutral (gray option) response since no criterion was ever truly neutral, (i.e. distribution of positive/negative was always 4-2, or 2-4, never 3-3.)

The distribution of ratings for the overall assessment is shown in Figure 7. Extreme scores were never selected. In the Text group, for six of the nine participants, the overall rating was neutral. In contrast, only two IOB participants rated overall as neutral.

Time

Response times were captured for (1) assigning all 30 items to a criterion and (2) scoring the criteria and overall assessment. Great caution has to be used in interpreting response time since the experiment was conducted on the internet and there was no experimenter present. The participants could have been interrupted by a phone call or a friend, stopped for a snack, etc., and the clock would still keep running. Greater confidence can be attached to the time to score for the five criteria plus the overall rating since these were relatively short, and therefore, less likely to have been interrupted.

Figure 8 shows the average response times to assign the first 30 items. As expected, the IOB group took significantly longer ($F_{1,14} = 8.24$, $p<.05$). The IOB group had the additional tasks of rating the effect of each information item. On the average, these extra tasks took 18 more seconds per item. Surprisingly, there was also a significant difference between the Positive and Negative decision groups ($F_{1,14} = 4.62$, $p<.05$), with participants in the Positive group taking longer to complete the 30 items. This was especially surprising for the Text group who were not required to formally rate the item (as was the case with the IOB group).

Figure 9 shows the average times required to complete scoring the five criteria and assigning an overall assessment. There was no significant Display effect: the use of IOBs did not effect decision making time. The average time to complete the six ratings was about 3.75 minutes. There was again a significant Decision group effect ($F_{1, 14} = 5.49$, $p<.05$) with participants in the Positive group taking longer to complete the six decisions (4.7 minutes versus

2.6 minutes) than those in the Negative group. Making a Positive rating appears to require more cognitive processing time than making a negative rating.

Errors

If a criterion had a predominance of positive items and the criterion was scored as negative by the participant, it is a scoring error. If the participant scored a criterion as neutral, this was also scored an error, but the situation is more ambiguous, since different weightings may have been subjectively assigned to the positive and negative items by the user.

Correctness was also determined by the items that the participant assigned to a criterion. For example, if a pre-experimental assignment was 4 positive and 2 negative items to the TSP criterion, but a Text participant's assignment had 5 items to the TSP, 3 of which were negative and 2 positive, the "correct" scoring would be Negative even though the pre-experimental classification was positive. This occurred in only a limited number of cases.

The IOB group assigned four effect levels to an information item (Red or Yellow for Negative, Light Green and Dark Green for Positive). "Correctness" in scoring a criterion was determined by the algebraic sum of the individual items (i.e. -2 for Red. -1 for Yellow, +1 for light Green, +2 for dark Green). Thus, if an IOB participant has a criterion with 5 items, two of which were assigned Red scores and three of which were light Green, the summary score would be -1 and the correct decision would be a negative assignment (Yellow or Red) of the criterion. If the criterion had 5 items, 1 Red, 2 Yellow, and 2 dark Green, the sum is zero and the correct assignment is Neutral (gray).

Figure 10 shows the error rate for scoring the five individual criteria. A criterion could be scored as "very negative", "negative", "neutral", "positive", or "very positive." The maximum number of errors is 45 each for the Text and IOB groups (9 participants with five criteria). The two columns on the left compare TEXT and IOB conditions, where Neutral is scored as an error (unless, of course, the item ratings summed to neutral). The rates are about the same (42% for Text, 37% for IOB). Both of these error rates are significantly lower than the 67% error rate predicted by chance guessing (i.e., participants item scores could sum to either positive, neutral, or negative). The two columns on the right show only the reversal errors (i.e., scoring a negative criterion as positive or vice versa.), ignoring those errors for a neutral score. This is certainly a far more serious error than assigning a wrong neutral score. The Text condition had six reversals (a 13% error rate) while the IOB condition had only one reversal (a 2% error rate).

The error rate for the overall assignment is important because it shows the basic 'bottom line" decision, (i.e., "after seeing all the data, what is your final recommendation?") If we score neutral as an error (remember the a priori sum of all the criteria were either positive or negative), we find, astonishingly, that not one of the nine members of the Text group made the correct overall rating. In contrast, six of the nine IOB members made the correct overall rating. Figure 11 shows the distribution of errors and indicates that the Text group had 6 members that selected the Neutral rating and three that wrongly assigned a Positive or Negative rating. These three were reversal errors (i.e. selecting positive instead of negative or vice versa.). Only two of the IOB participants selected neutral as the overall rating. Only one of the nine IOB participants made a reversal error, the other six made correct decisions.

IOB usefulness

At the conclusion of the experiment the IOB participants were ask to rate the usefulness of having IOBs available when they made the criterion and overall ratings. Of the nine participants, three rated the IOBs as "somewhat helpful", and six rated them as being "very helpful".

Discussion

There are several major findings in this experiment. Participants, who only had text items without IOBs during the final decision, made poor overall assessments. Although the IOB participants performed better in making the overall assessment, they did not perform better than the Text participants in making the individual assessments for each of the five decision criteria. Finally, both groups took longer to process information when rating the information as positive and supportive rather than negative and critical. Each of these areas is discussed below.

Not a single participant in the Text condition made the correct overall recommendation. They tended to select "Neutral" even though there was always a preponderance of information in favor of either a positive or negative judgment. The three participants in this Text condition who did not select Neutral made incorrect final recommendations (a reversal). The IOB participants performed much better, with six of the nine making the correct overall assessment. Only one participant in the IOB group made a reversal error.

The Text group tended to select the Neutral option in the overall assessment even though Neutral was not a correct composite assessment of how they had scored the criteria. There are several possible reasons for this finding. One is that these participants weighted the criteria differently, even though they had been instructed that each criterion was of equal value. If they had two negative criteria and three positive criteria and put more weight on the negative criteria, this would result in an overall neutral rating. This was unlikely to have happened, because the IOB participants did not show a strong preference for the Neutral option. A more likely reason is that the Text participants did not have very much confidence in the accuracy of their recommendation and selected Neutral as a safe, "middle of the road" assessment. Text participants had difficulty keeping track of the subjective assessments of the 30 different information items and assigning them to five distinct decision criteria. It was further required to mentally combine all the text items for each criterion, and assess each criterion. Then had to review all five criteria scores, remember how they were determined, and finally make a single overall assessment. The easiest resolution to this task was to make a Neutral overall assessment regarding the composite picture created by the 30 text items.

The IOB participants, on the other hand, had to intuitively combine sorted color bars for each item, a much easier "search and assess" cognitive task. When they completed the scoring of the five criteria they were confident that they had correctly processed all the information and were willing to decide on something other than the Neutral option as the overall assessment. The confidence of a particular rating score by treatment group will be explored in our follow-on work.

If the Text participants did not have the Neutral option available, would they have made the correct assessment? Had the Text participants incorrectly processed the information or had they correctly processed the information, but were simply not willing to commit to a positive-negative rating because of lack of confidence in their decision? It may be desirable in the next experiment to give all subjects the option of selecting a Neutral rating, but if they chose that option they will be forced to make a provisional positive-negative choice. This may provide us some insight regarding whether they would have made the correct choice had the Neutral option not been available.

A basic assumption in the use of IOBs is that IOBs become increasingly beneficial as the cognitive workload of the assessment task increases. This may be the reason why there were no performance differences between the Text and IOB participants for scoring the individual criterion. For each criterion, each participant had to only assess the relative contribution of a small number of information items, a simple task that did not benefit from a list of IOBs. In contrast, to derive an overall assessment the participants had to review the scoring for each of the five criteria and come to a composite assessment. IOBs seemed to help with this cognitive process.

The significant increase in time to process a preponderance of positive versus negative items is somewhat surprising, especially for assigning the 30 items to a criterion. A case could be made that the IOB participants (who had to rate the supportiveness of each item) might be slower to commit to a positive assessment, since it helps endorse an action (using Islandia) whereas a negative scoring supports the more conservative response of no action (i.e., no, don't use Islandia). This reason is unlikely because the Text participants also took a long time to assign the 30 items when the preponderance of the items were positive in nature. It may be that people can more quickly assess a negative impact item ("This is not good."), while a more positive item needs more attention and more careful mental review ("Is this really supportive?").

A similar significant time difference was found for the time to score the five criteria and rate the overall assessment. Once again, the contributing factor may be the participants desire to be more certain when they are recommending an action and committing resources. It may also be that some participants found a "fatal" negative item (i.e., a "show stopper"), an important item that does not allow the participant to recommend action. Verbal reports from another study (Fleming & Cowen, in prep.) suggest that this may be the case for some participants, enabling them to make a quick negative recommendation.

In summary, we have collected some data on three important issues about the value of iconic abstractions of subjective knowledge: (1) Does the abstraction process create an undue cognitive burden on the individual? We found that processing time to evaluate the effect of the item and type in a key word tag was only about 18 seconds. (2) Do users like using the IOBs in their decision process? We found that none of the nine IOB participants negatively rated the value of IOBs and six selected the most favorable rating. (3) Do IOBs improve decision quality above and beyond the decisions made with original text information? In this study, IOBs had no effect upon decision time or the quality of decision making when rating criteria. However, the IOBs participants performed better that those in the Text group in making the overall assessment. This implies that IOBs will increase one's confidence in the information building process and

makes an individual more willing to commit to the right go/no-go decision. Our results lend support to the belief that IOBs can improve individual decision making performance for a relatively simple task scenario. Future studies will increase the complexity of the task and will involve the sharing of the IOBs among distributed participants.

References

Dennis, A. R. (1996). Information Exchange and Use in Group Decision Making: You Can Lead a Group to Information But You Can't Make It Think. MIS Quarterly, vol. 20, no. 4, pp. 433-455.

Fleming, R. and Kaiwi, J. The Problem of Unshared Information in Group Decision-Making: A Summary of Research and a Discussion of Implications for Command and Control. SPAWAR Systems Center-San Diego (SSC-SD) Technical Document 3149, Dec., 2002

Fleming, R. Information Exchange and Display in Asynchronous C2 Group Decision Making. Paper to be presented at 8th International Command and Control Research and Technology Symposium, Washington D.C., June 2003.

Fleming, R. and Cowen, M. The Effects of New Information on Decision Modification: The Role of Display Format and Direction of Impact. (SSC-SD Technical Report, in preparation.

Wittenbaum, G. M., and G. Stasser (1996). Management of Information in Small Groups. In Nye, J. L. and A. M. Brower (Eds.). What's Social About Social Cognition? Research on Socially Shared Cognition in Small Groups. Sage, Thousand Oaks, CA, pp. 3-28.

Wittenbaum, G. M. (1998). Information Sampling in Decision-Making Groups: The Impact of Members' Task-Relevant Status. Small Group Research, vol. 29, pp. 57-84.

Table 1. Overall Experimental Design.

		Display		
		Text Only	IOBs	
Decision	Positive	N=5	N=5	N=10
	Negative	N=4	N=4	N=8
		N=9	N=9	**N (tot)=18**

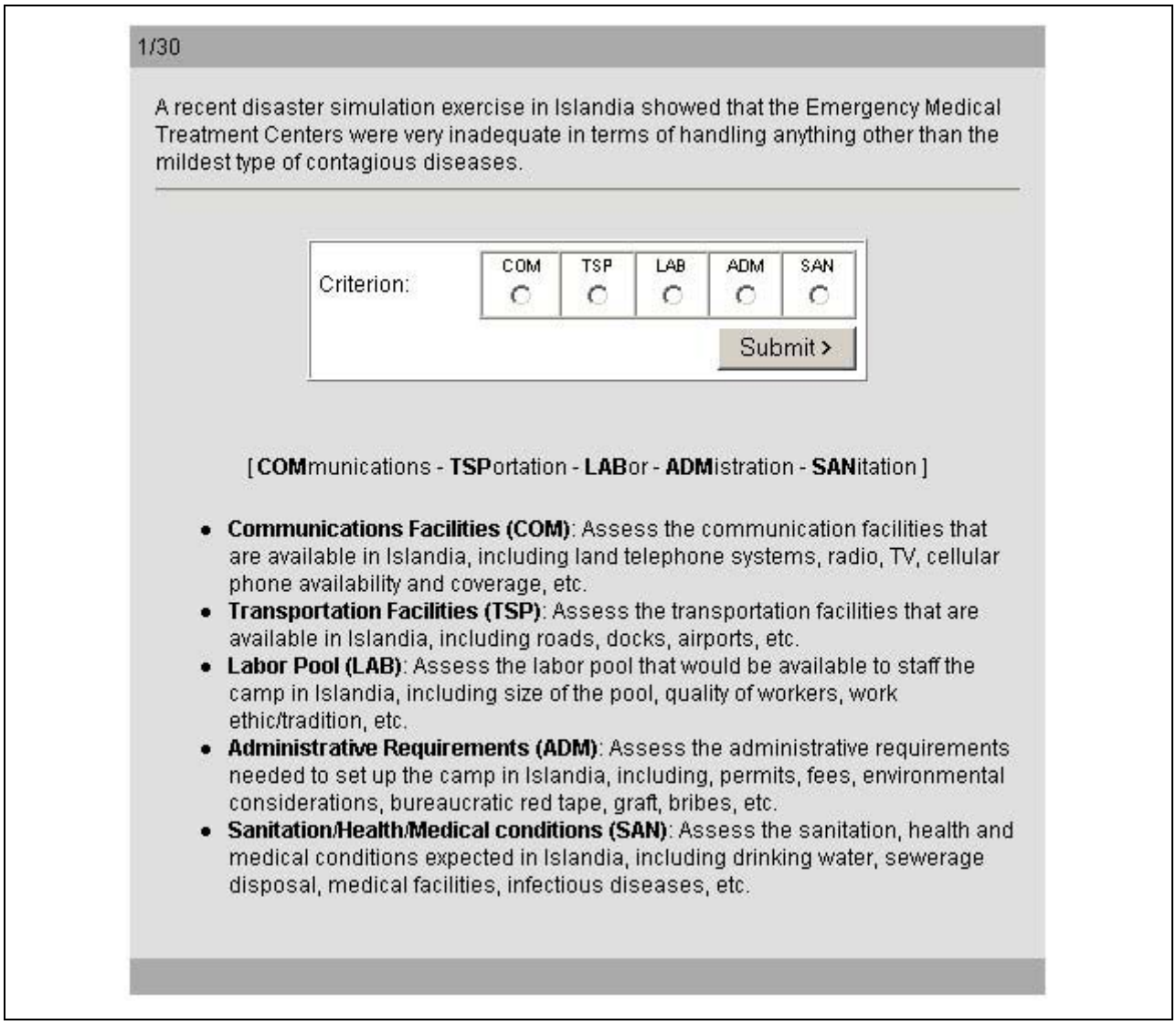


Figure 1. Example of the 30 information item classifications required by the Text group.

.

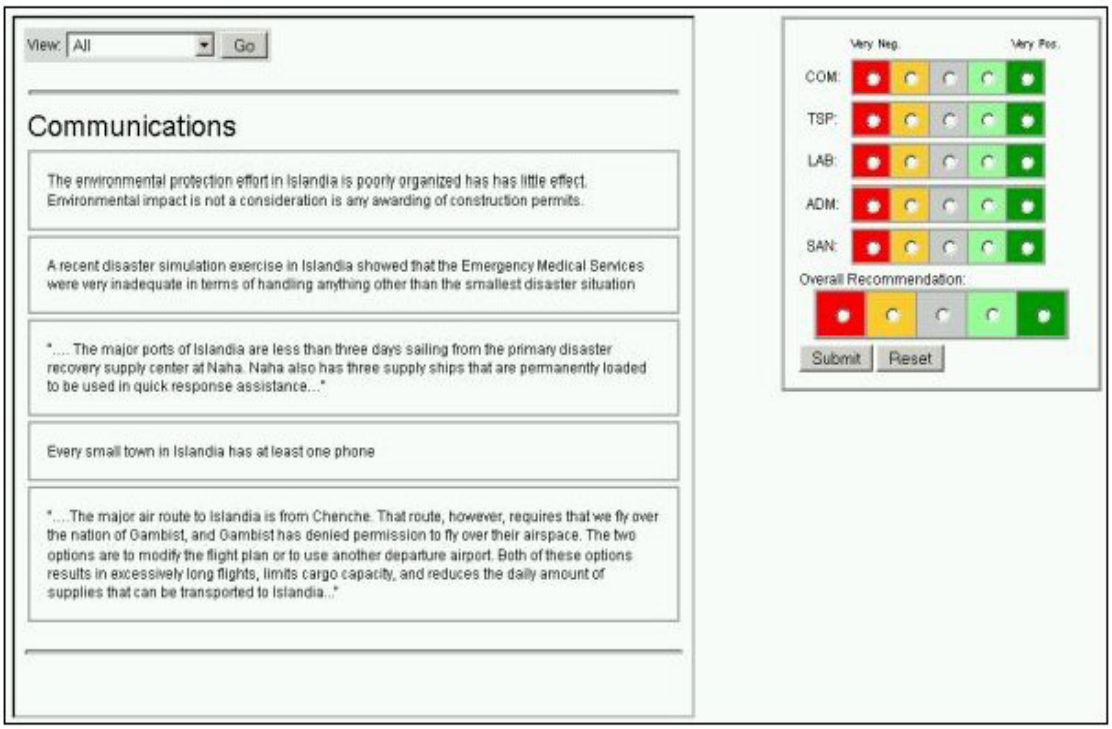
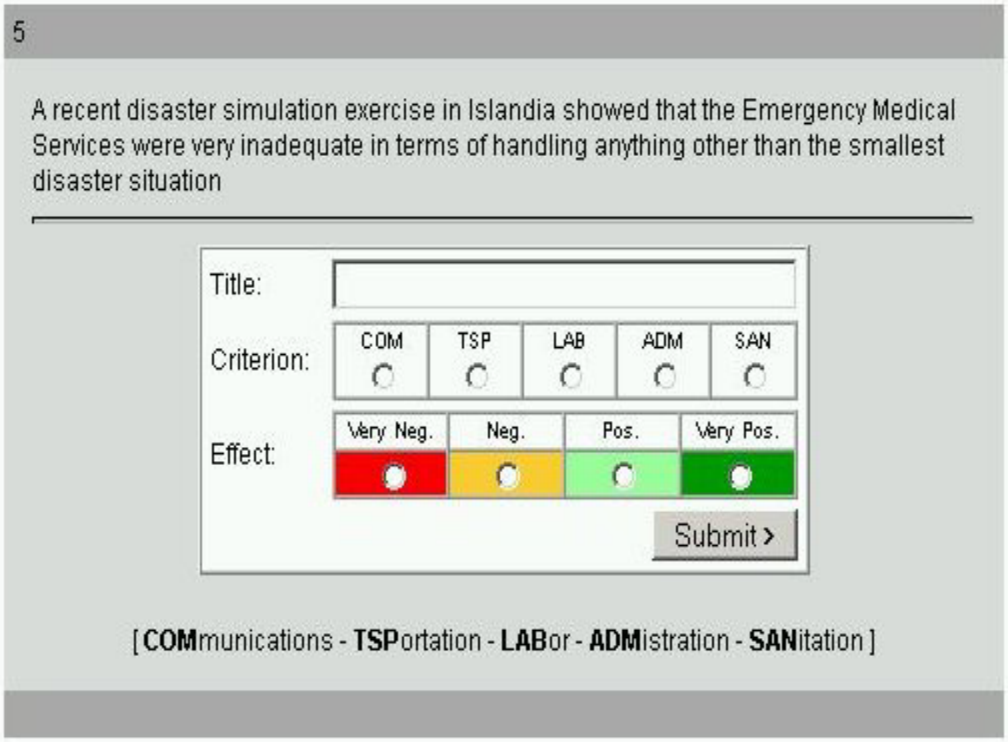


Figure 2. Example scoring page presented to Text group.

Here is a sample of the kind information items you will see.



After you read the item, you will:

1. enter a **short keyword** identifier in the Title box to help you recognize the item later
2. assign it to what you feel is the most **appropriate criterion** and
3. **evaluate the effect** of the item on that criterion by assigning it a rating on a four-point scale. The four ratings are "Very Negative", "Negative", "Positive" and "Very Positive".
    - For example a road construction delay of one-week might be assigned to the Transportation criterion and rated as "Negative".
    - If the delay was reported as three months instead on one week, it might be rated as "Very Negative".

When you are satisfied with your scoring of the item, click "Submit" to advance to the next item.

Figure 3. Page 2 of instructions given to IOB group.

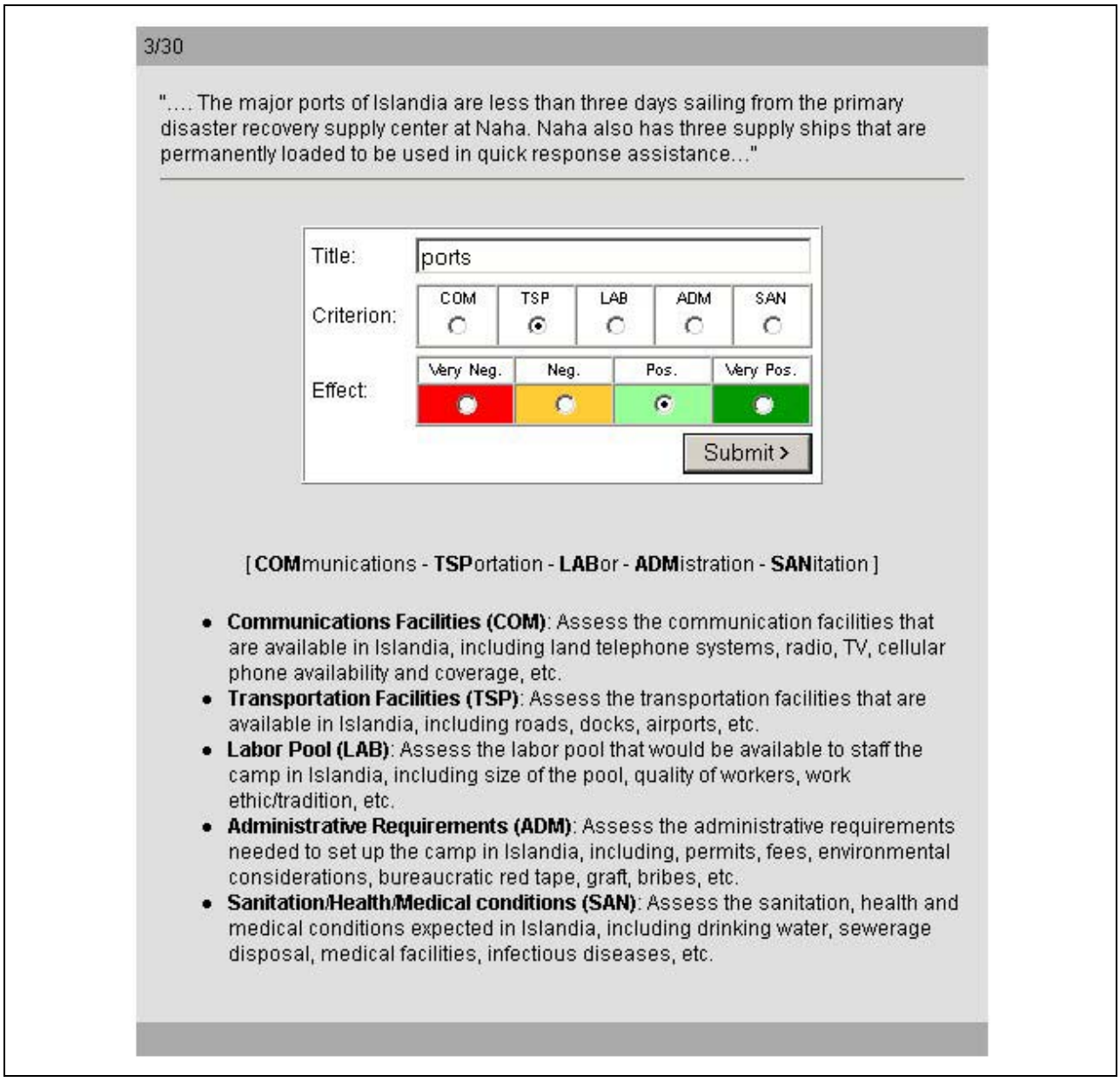


Figure 4. Example of an item that has been scored by a member of the IOB group.

On the right side of this final display, you will see the rating box to be used for the five criteria and the overall assignment. Review the IOBs associated with a criterion, and then assign your rating to this criterion by clicking the most appropriate rating button for that criterion.

When you have completed rating the five criteria, review these criterion assignments and make a final overall recommendation as to the advisability of using Islandia. You can change your ratings at anytime up until you click the "Submit" button, which will lock in the ratings.

Below is a SAMPLE of the scoring page. To get to your items, click the 'advance to scoring' button.

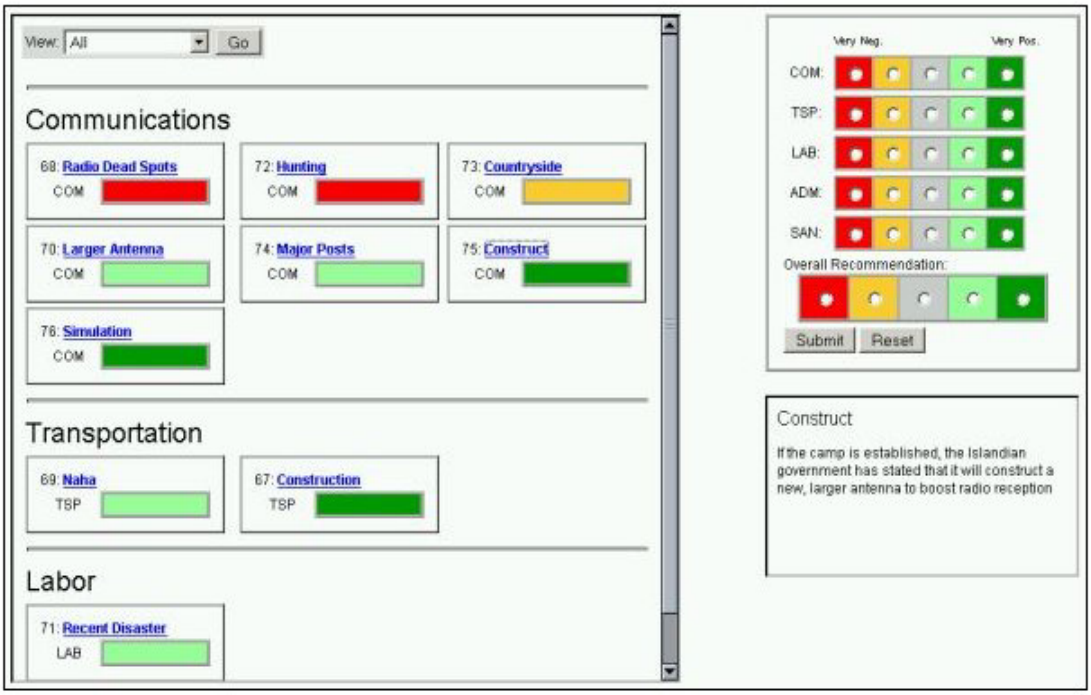


Figure 5. Page 2 of instructions given to the IOB group after they completed scoring their 30 items.

.

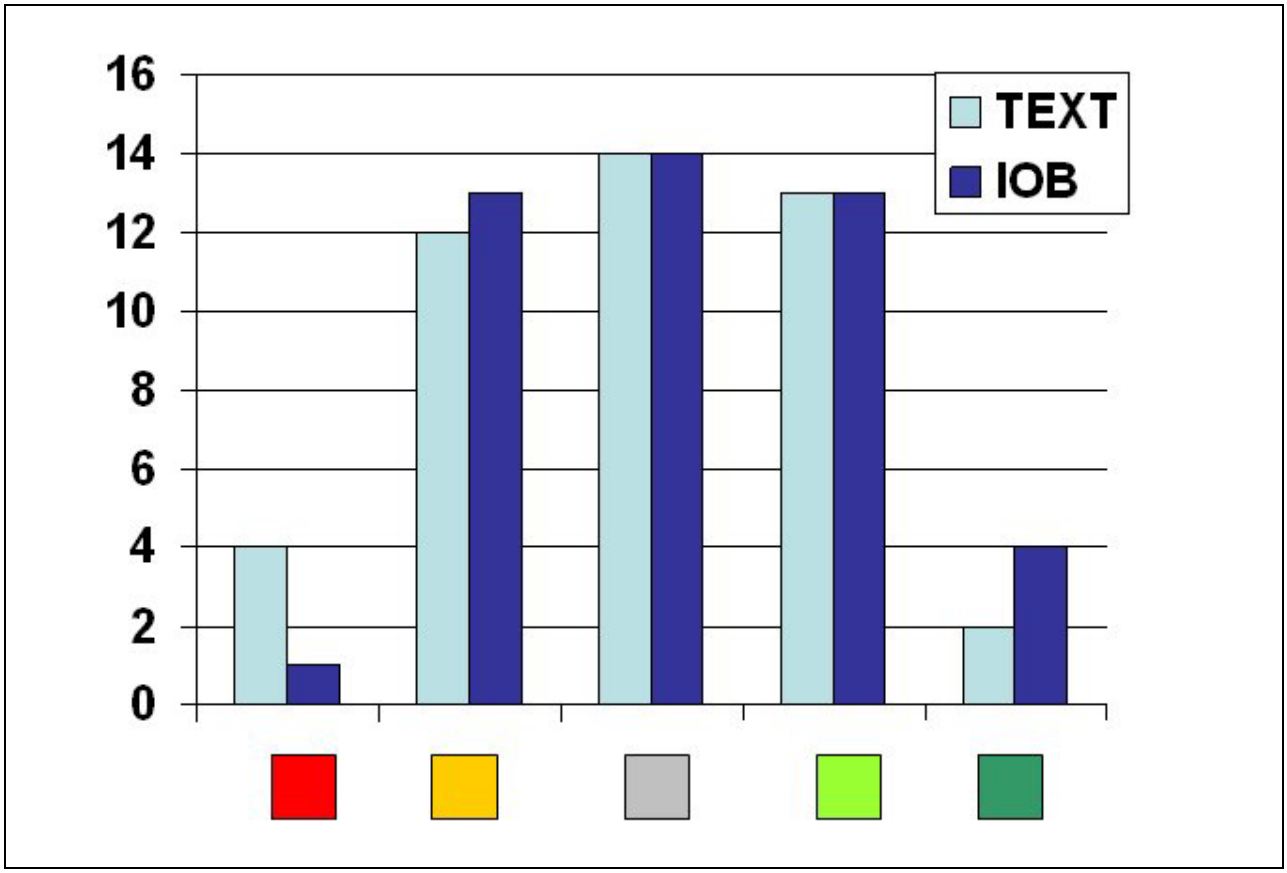


Figure 6. Frequency of use of the 5 scoring options when making criteria decisions.

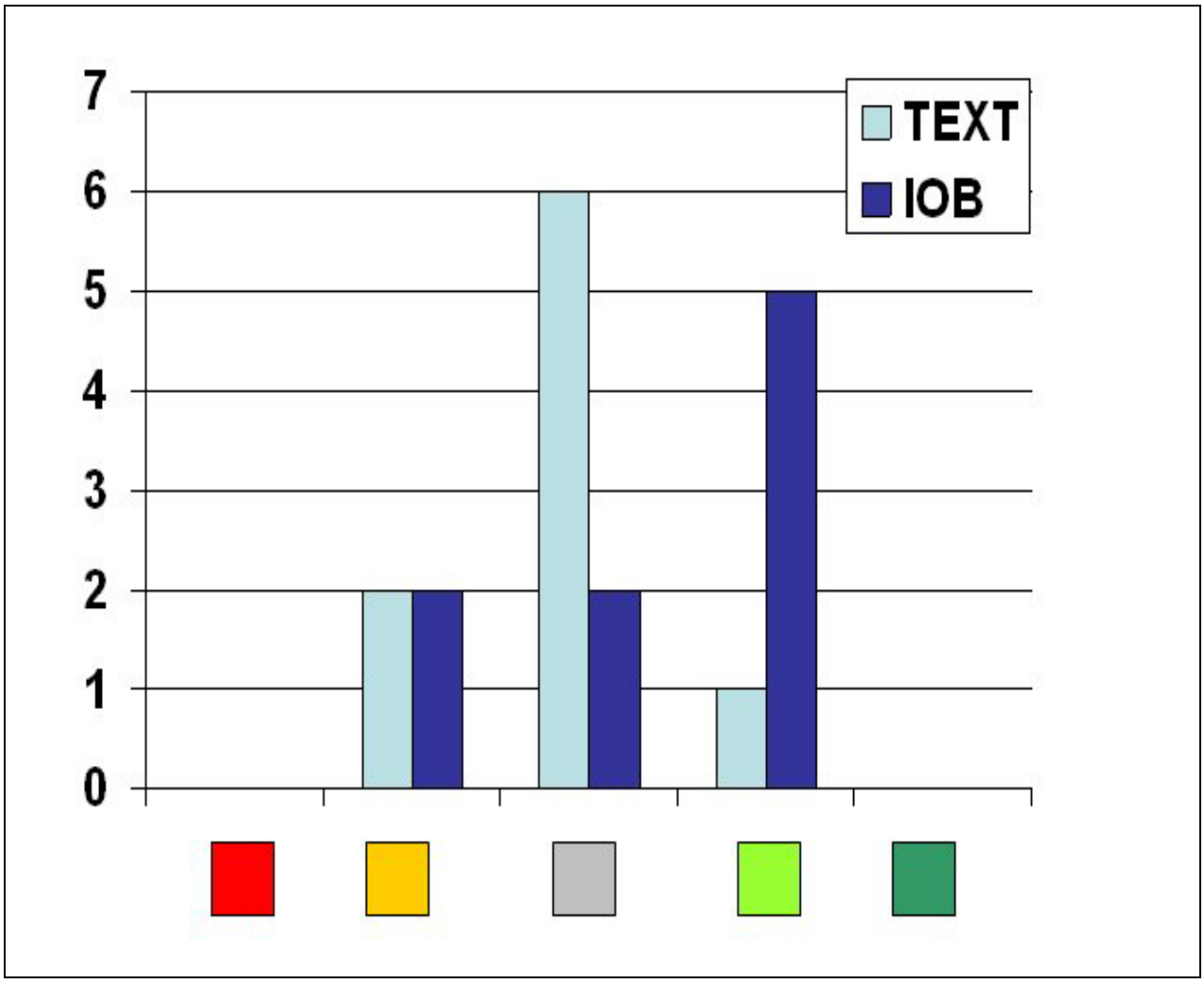


Figure 7. Frequency of use of the 5 scoring options when making the overall decision.

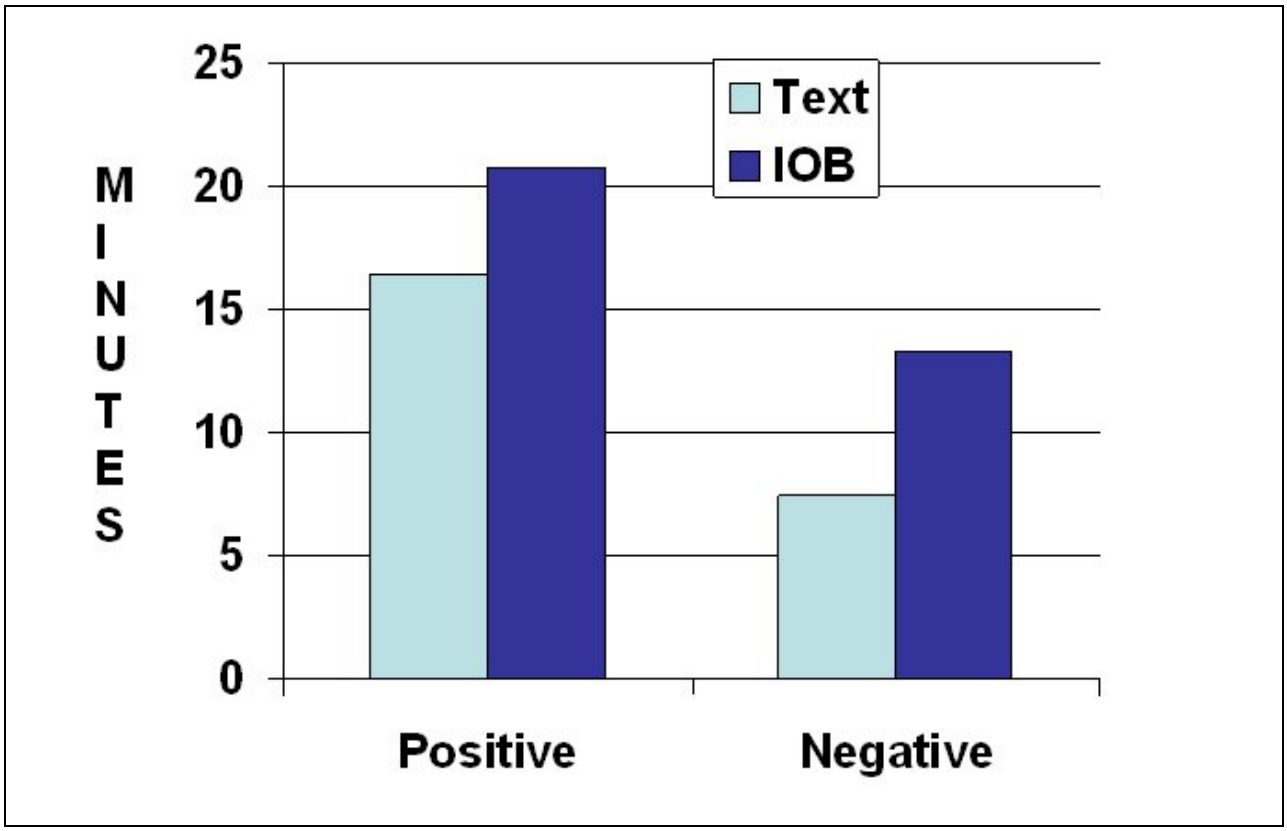


Figure 8. Average time required to assign the 30 items.

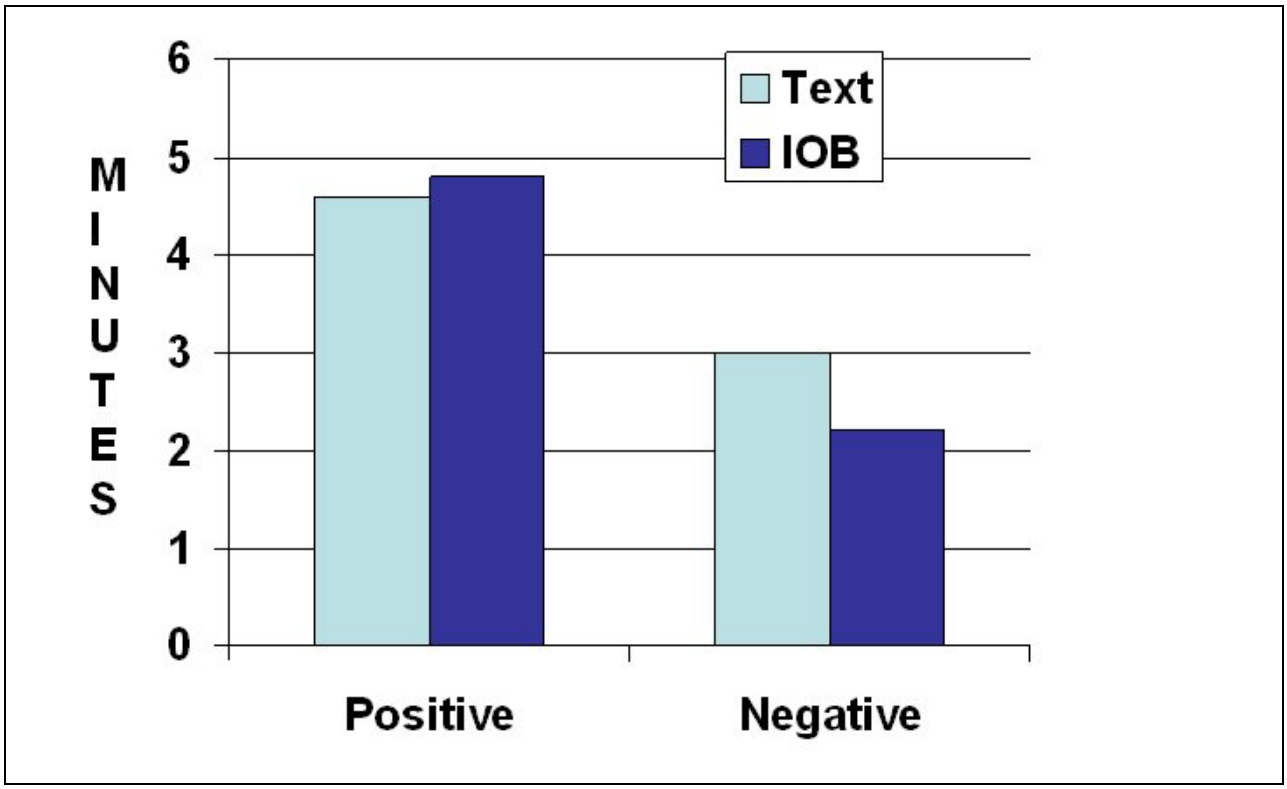


Figure 9. Average time required to complete scoring the five criteria and the Overall score as a function of whether the information was predominately Positive or Negative.

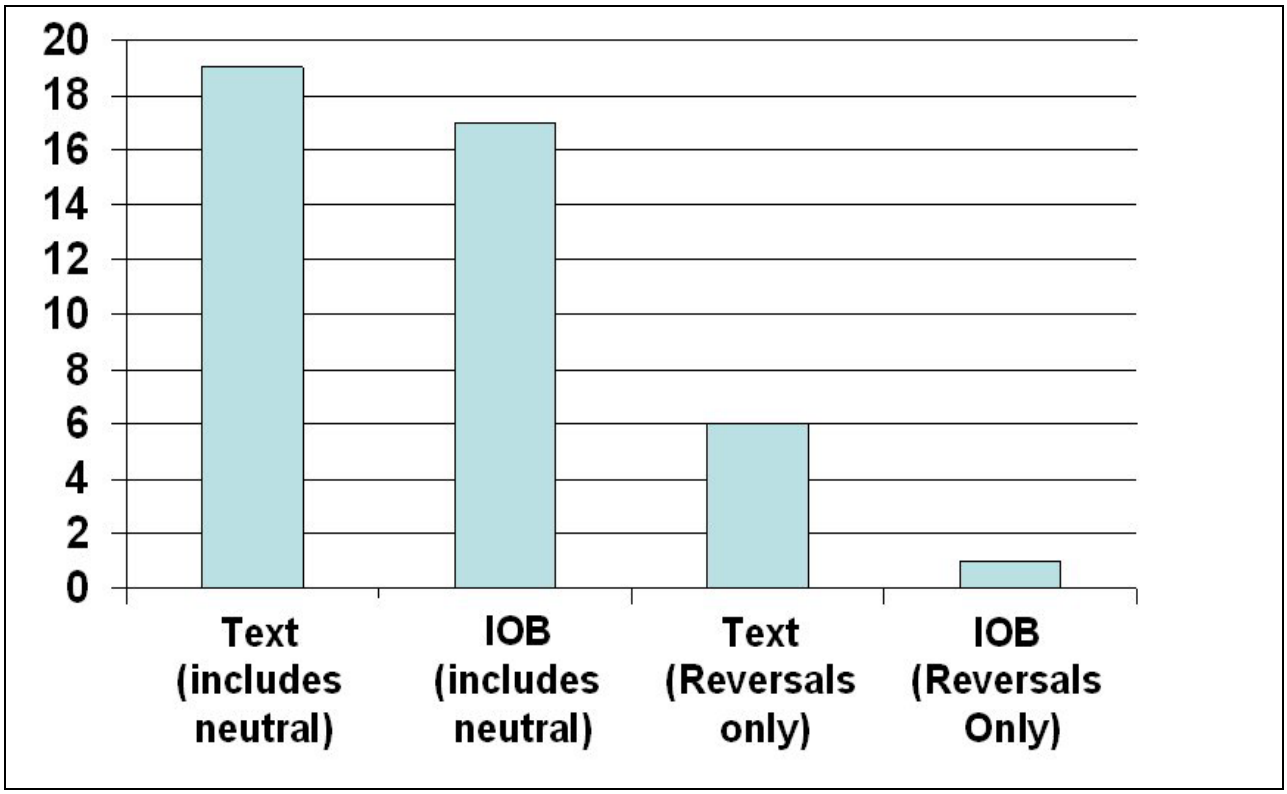


Figure 10. Number of scoring errors while scoring the five criteria ("Reversals Only" do not count an incorrect neutral score as an error).

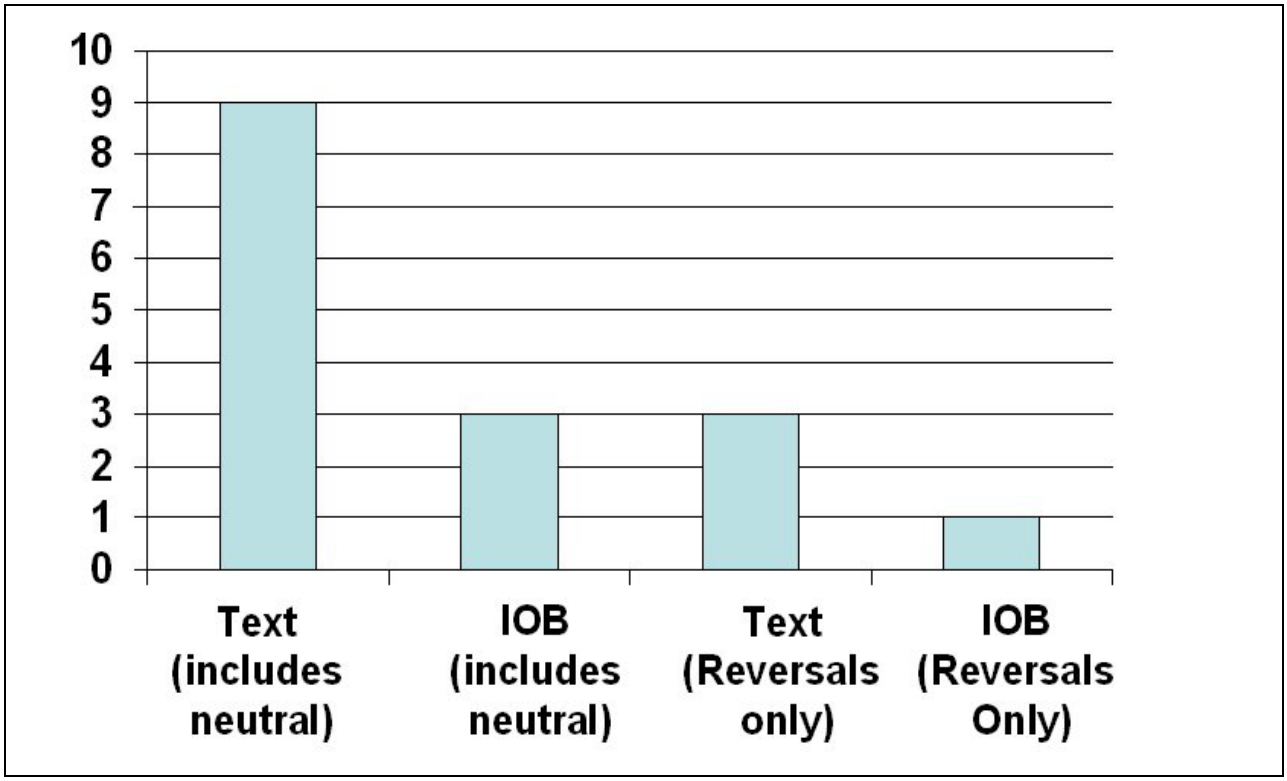


Figure 11. Distribution of error types for assigning overall scores.