Technical Report 1070

Scoring System Improvements to Three Leadership Predictors

Michelle R. Dela Rosa
Human Resources Research Organization

Deirdre J. Knapp
Human Resources Research Organization

Brian D. Katz
Human Resources Research Organization

Stephanie C. Payne
George Mason University
Consortium Research Fellows Program

September 1997

United States Army Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Melanie Gorenflo-Gilbert, City University of New York
Dennis D. Stewart, Department of Behavioral Sciences
and Leadership, USMA

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED
1997, September	Final	10 July 1996 to 31 July 1997

4. TITLE AND SUBTITLE
Scoring System Improvements to Three Leadership Predictors

5a. CONTRACT OR GRANT NUMBER
MDA903-93-D-0032 DO0048

5b. PROGRAM ELEMENT NUMBER
A790

6. AUTHOR(S)
Michelle R. Dela Rosa, Deidre J. Knapp, Brian D. Katz

(HumRRO); and Stephanie C. Payne (George Mason Univ.)

5c. PROJECT NUMBER
20262785A790

5d. TASK NUMBER
1115

5e. WORK UNIT NUMBER
C01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Resources Research Organization
66 Canal Center Plaza, Suite 400
Alexandria, Virginia 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
DFR-EADD-97-19

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
Attn: PERI-RM
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. MONITOR ACRONYM
ARI

11. MONITOR REPORT NUMBER
Technical Report 1070

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
COR: Trueman R. Tremble, Jr.

14. ABSTRACT: This project sought to examine and improve the reliability of the scoring systems for three instruments which have been used in previous Army leadership research. Review of existing literature and interviews with project staff participating in prior research provided initial information concerning the strengths and weaknesses of the scoring systems for the three instruments. This information was used to recommend modifications to the original scoring systems. Six individuals were trained to use the modified scoring systems. The modified scoring systems were evaluated by rescoring responses randomly selected from the sample which had been scored according to the scoring systems originally developed for the leadership research program. Reliability estimates for the three modified scoring systems were consistently strong and showed improvements over those obtained through the original scoring systems. Interrater agreement indices were significant for nearly all ratings. Validity estimates provided evidence that each modified instrument was moderately to highly correlated with conceptually similar scores generated through the original scoring scheme. The report recommends use of the revised rating systems in future research to improve the quality of measurement from the three predictors.

15. SUBJECT TERMS
Leadership Open-Ended (Constructive) Tests Problem Solving

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	145	

Technical Report 1070

Scoring System Improvements to Three Leadership Predictors

Michelle R. Dela Rosa
Human Resources Research Organization

Deirdre J. Knapp
Human Resources Research Organization

Brian D. Katz
Human Resources Research Organization

Stephanie C. Payne
George Mason University
Consortium Research Fellows Program

Leader Development Research Unit
Trueman R. Tremble, Jr., Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

September 1997

Army Project Number
2O262785A790

Personnel Systems and Performance Technology

FOREWORD

The Center for Leadership and Organizations Research (CLOR), jointly established by the U.S. Military Academy (USMA) and U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) between 1992 and 1997, conducts programmatic research on Army-wide priorities in the areas of organizational leadership and of leadership education, training, and development. In 1994, CLOR initiated a program of longitudinal research to increase the understanding of the leadership development process. This program has involved the development of the Baseline Officer Longitudinal Data Base (BOLDS). Begun with cadets in USMA Class of 1998, completed BOLDS will contain data on leadership from several perspectives for describing changes in leadership behavior over time and for identifying experiences which contribute to successful leader development.

One current perspective defines *leadership* in terms of the solution of organizational problems. BOLDS will contain measures of problem-solving capabilities which are obtained by assigning scores to responses made to open-ended, written exercises. The problem-solving measures had originally been adapted for research use in an earlier ARI project on leadership at different organizational levels. The earlier project had shown promise for the measures.

The research reported here sought to ensure reliable measurement of problem-solving capabilities throughout the period of BOLDS data collection. Based on review of the problem-solving exercises, the researchers recommended and made modifications to the procedures guiding assignment of scores to the open-ended responses and to the training given to individuals serving as scorers. The researchers then assessed the modifications by comparing the reliability and validity of scores obtained through the modified and the original scoring systems. The modifications successfully improved scoring reliability. The improvements suggest the potential of the modified systems in longitudinal or comparative research on leadership problem solving and the development of this capability.

ZITA M. SIMUTIS
Technical Director

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

The authors would like to extend their appreciation to Michelle Marks and Cindy Parker for providing information on the original scoring systems and the archival data set. Special thanks are also extended to Dr. Trueman Tremble, Dr. Thomas Kane, Elizabeth McLaughlin, and Maggie Collins for scoring the instruments and providing valuable feedback.

SCORING SYSTEM IMPROVEMENTS TO THREE LEADERSHIP PREDICTORS

EXECUTIVE SUMMARY

Research Requirement:

Mumford, Zaccaro, Harding, Fleishman, and Reiter-Palmon (1993) developed a model of the components related to effective leadership. This model is the foundation of a research program designed to examine the relationships among cognitive ability, knowledge, temperament, and problem-solving skills and their prediction of effective Army leadership. In addition to other measures, three open-ended instruments, *Consequences, Alternate Headlines,* and *Military Scenarios*, have been used in this leadership research. The purpose of the present project was to examine and improve the reliability of the scoring systems for these instruments.

Procedure:

A review of the existing literature provided initial information concerning how the three measures were previously scored in ARI's leadership research program, and interviews with research staff provided their perspective of the strengths and weaknesses of these scoring systems. This information was used to make recommendations for modifications to the scoring systems. To evaluate these modifications, responses were randomly selected from a sample that had been previously scored using the scoring systems initially developed for the leadership research program. Six individuals were trained to score these responses using the modified scoring systems. Three individuals rated each response. Descriptive statistics were computed for each instrument's individual scales, as well as for composite scores. Reliability analyses were conducted for four data sets which were created for each instrument. The validity of the instruments was also evaluated by examining the relationship between scores on the three instruments and scores on other related measures. These analyses were used to make final recommendations concerning the scoring of these instruments.

Findings:

Using the modified scoring systems, the reliability estimates for the three modified instruments were consistently strong and showed an improvement over those generated with the original scoring system. This improvement was particularly strong for the *Military Scenarios* and *Alternate Headlines* instruments. The reliability estimates for the measures were quite consistent and strong with three scorers, but the estimates calculated using two scorers had more variability depending on the specific scorers included in the calculation. Furthermore, interrater agreement indices were significant for nearly all *Alternate Headlines* and *Military Scenarios* ratings (interrater agreement indices were not possible for the *Consequences* instrument).

Validity estimates provided evidence that the Demonstration scores were moderately to highly correlated with conceptually similar scores from the Archival data set. Though the Archival scores were more highly correlated with external measures, this apparent advantage over the Demonstration scores was largely reduced when correlations that partialed out the influence of rank were computed. Unlike the Demonstration scorers, the Archival data set scorers knew the rank of respondents, and this could have influenced their ratings.

Despite the overall success of the modified scoring systems, the *Alternate Headlines* Presentation scale requires further modification. The scores generated using this scale were restricted in range because most scorers failed to use the entire scale. This consequently impacted the scale's reliability. As a result, an alternate presentation scale is recommended.

Utilization of Findings:

The modifications made to the scoring systems successfully improved each instrument's reliability. The modified scoring systems are recommended for use in scoring data that has been collected from Academy cadets at West Point and for future uses of these instruments.

SCORING SYSTEM IMPROVEMENTS TO THREE LEADERSHIP PREDICTORS

CONTENTS

Page

Appendices

List of Tables

Scoring System Improvements to Three Leadership Predictors

Introduction

The U.S. Army Research Institute for Behavioral and Social Sciences (ARI) sponsored a multiple-study leadership research effort to develop and test a model of organizational leadership. Forming the foundation for this research program was a model developed by Mumford and his colleagues (Mumford, Zaccaro, Harding, Fleishman, & Reiter-Palmon, 1993) that defines organizational leadership as "discretionary problem solving in ill-defined domains" (p. vii). Their leadership model uses cognitive and temperament variables to predict organizational leadership.

Mumford et al. conducted two studies to test the problem solving leadership model, one using Army officers and the other using civilian managers (MRI, 1995). ARI researchers attempted a partial replication of the Mumford et al. findings using data collected from another sample of Army officers (Tremble, Kane, & Stewart, 1997). ARI has also collected data on West Point cadets with the intention of conducting a longitudinal research study to examine the hypothesized leadership model.

These studies have used both standard and newly-developed open-ended instruments to tap cognitive and problem solving capabilities. These instruments present research participants with a scenario or other type of stem which prompts them to generate or construct open-ended responses. Scoring rules are then applied to derive scores on these open-ended measures.

Purpose

The purpose of this effort was to improve the reliability of the scoring systems for three open-ended instruments that have been used in the ARI problem solving leadership research program. In the process of addressing reliability concerns, there was also considerable attention given to improving the construct validity of the scoring schemes. The three instruments are *Consequences, Alternate Headlines*, and *Military Scenarios*. *Consequences* has been used as a standardized measure of creative thinking (Guilford & Guilford, 1980). The second measure, *Alternate Headlines,* taps creativity and has also been used as a measure of writing skill. The third instrument, *Military Scenarios*, was developed specifically for this leadership research program to measure how well individuals define and analyze problems.

Approach

The basic methodology for the present study was to propose changes to the scoring strategies for the three instruments and demonstrate whether these changes improved upon the original scoring systems by re-scoring responses collected in the initial model validation study (MRI, 1995). Project staff began by reviewing the studies conducted as part of the problem solving leadership research effort and conducting interviews with researchers who had participated in these studies. Based on this information, recommendations were made regarding

modifications to the original scoring systems. A demonstration study was designed and implemented to compare the modified scoring systems to the original systems by statistically analyzing a sample of ratings using both scoring systems. In addition, feedback was collected from scorers regarding the ease with which the new scales were used.

Organization of the Report

This report is organized as follows. The next section describes the research that has been conducted as part of the ARI problem solving leadership research effort and reports on the results of interviews with researchers involved in that work. The following section describes the scoring system modifications that were recommended for the present study. The design of the demonstration study plan, scorer training, and the procedures used to collect ratings on the instruments are outlined in the section after that. This is followed by a description of the demonstration study results. Final recommendations for further use and scoring of the target measures are presented in the last section of the report.

Background on the Research Program

To help assure that scoring system recommendations made in the present project were consistent with the goals of the Mumford et al. leadership model (1993), project staff reviewed this model and the validation research associated with it. This helped provide an understanding of what the instruments were originally intended to measure, how the instruments performed statistically in this research, and how the instruments correlated with other research variables of interest. Interviews with individuals who had served as scorers in the model validation studies were helpful for understanding some of the problems they experienced in using the scoring systems devised by Mumford et al.

Problem Solving Leadership Model Research

Mumford et al. (1993) developed an executive leadership model that defined leadership as "discretionary social problem solving in ill-defined domains" (p. 9). Their model specifies a taxonomy of 13 Leadership Behavior Dimensions organized into four superordinate dimensions (information search and structuring, information use in problem solving, managing personnel resources, and managing material resources). The Leadership Behavior Dimensions encompass sets of tasks normally associated with leadership positions. Mumford et al. also hypothesized a taxonomy of 65 leadership knowledge, skills, abilities, and personality constructs (KSAPs). The KSAPs were organized into four dimensions (cognitive generating factors, personality, values and motives, and embedded appraisal and implementation skills) and 11 subdimensions.

Alternate Headlines was used as a measure of crystallized cognitive skills (those associated with writing skill) and the *Consequences* instrument was used as a measure of creativity. Both crystallized cognitive skills and creativity are two of the 11 KSAP subdimensions proposed by the Mumford et al. model. It is unclear whether *Military Scenarios*

was intended as a measure of solution construction skills, one of the 65 KSAPs proposed by the model, or as a measure of a problem solving construct that appears later in the model.

Mumford and his colleagues conducted studies using military and civilian samples to validate their leadership model. The "Officer study" sampled over 1800 O1 (2nd Lieutenant) to O6 (Colonel) Army officers. The results of this study were reported in a briefing by the Management Research Institute (MRI; 1995). The Officer study showed preliminary supporting evidence for the leadership model, particularly the contribution of the complex problem solving skills to leader effectiveness. Interrater reliability estimates for the scales associated with the three target instruments ranged from a high of .82 (*Consequences*) to a low of .57 (*Alternate Headlines*).

These same researchers also collected data from 543 civilians in a study they referred to as the "Civilian study." Data were collected on all three instruments in the Officer study, but a lack of time prohibited the collection of *Alternate Headlines* responses in the Civilian study. The results of the Civilian study have not been formally reported.

In an effort to replicate findings in the Mumford et al. Officer study, Tremble, Kane, and Stewart (1997) tested a part of the leadership model using a sample of officers from eight Army posts. This research provided limited support for the leadership model.

Longitudinal data have been collected from cadets at West Point, and the plan at the start of this project was to score these cadet responses using the scoring systems derived from the present study. At the time of this writing, however, plans for collecting additional data from cadets and following these individuals at key points in their military careers have been suspended.

Interviews

Interviews were conducted with researchers who scored the instruments in these previous studies. The researchers described the difficulties they had using the instrument scoring scales and concerns they had about the instruments. Researchers participating in the Civilian study reported that scorer training had been relatively unstructured. It had primarily consisted of a general discussion about what the scales meant and practice scoring a few responses. The researchers expressed some concern that their ratings were overly subjective and that as a group their interpretation and understanding of the scoring scales was not uniform. The interviews highlighted the need for a thorough, standardized scorer training workshop to provide instructions for using the scoring scales.

The interviews also verified the need to revise the existing scoring scales. Although the scales used for the leadership research were developed to tap multiple dimensions of leadership ability, the scorers indicated that they believed that several of the existing scales were redundant. For example, *Military Scenarios* was designed with as many as eight scales that could be used to

rate responses; however, scorers felt that there were at best two dimensions retrievable from the data, and that several of the dimensions were potentially providing redundant information (e.g., "quality" was really a global score including all other scores).

Recommended Changes

Based on the literature review and interviews, recommendations were developed regarding modifications to the three open-ended instruments, *Consequences*, *Alternate Headlines*, and *Military Scenarios.* The recommendations focused on scoring and administration of the three instruments. In the following section, scoring system recommendations specific to each instrument are presented. Administrative issues that apply to all three instruments follow these recommendations. Because the plan was to use the instruments in a longitudinal research effort that had already began, only cosmetic changes to the instruments themselves were considered.

Scoring System Modifications

A copy of each instrument is provided in Appendix A. Described below is a review of the scales proposed by original authors (where applicable), scales that were used during the leadership research studies, and recommendations regarding modifications to the scales.

Consequences

The *Consequences* measure developed by Guilford and Guilford (1980) presents five hypothetical situations with an open-ended response format so that examinees can report possible consequences of each situation (see Appendix A). The original scoring system is rooted theoretically in the Structure-of-Intellect model (Guilford & Hoepfner, 1966) and provides a standardized measure of creativity. The Guilford and Guilford scoring system results in two separate scores, Originality and Ideational Fluency. The Originality score is the sum of the acceptable, remote responses. The Ideational Fluency score is calculated as the sum of the acceptable, obvious responses. Responses are considered remote to the extent to which they are novel and imaginative and to the extent that they differ from the material presented in the problem. Obvious responses are consequences that directly result from the situation presented. The scoring scheme used in the leadership research studies, however, differed substantially from that created by Guilford and Guilford.

A total of eight *Consequence* scales were developed and used in the ARI leadership studies. However, all eight scales were used in only one data collection, the Officer study. The eight dimensions were Quality, Originality, Realism, Time Span, Negative Outcome Sensitivity, Positive Outcome Sensitivity, Complexity, and Degree of Abstraction. In the Civilian study, ratings were completed only on the Quality, Originality, and Realism scales. Regardless of how many of the *Consequence* scales were rated, a measure of creative thinking capacity was

calculated by averaging the ratings of quality and originality across scorers. Table 1 shows the Quality, Originality, and Realism scales that were rated in all of the leadership research studies.

HumRRO recommended a return to the Guilford and Guilford (1980) scoring system. Using this scoring method, Originality and Ideational Fluency scores are based on the number of obvious and remote responses generated, not upon an overall rating of the consequences presented for the situation. In addition, the development of the Guilford and Guilford scoring system was based on a clear theoretical framework and has been supported and validated by prior research.

Table 1. Original Consequences Scales

Quality:	**Originality:**	**Realism:**
How coherent, meaningful, and logical are the consequences with respect to the questions being asked? 1. Low quality 2. Below average quality 3. Average quality 4. Above average quality 5. High quality	To what degree are the consequences novel and imaginative? To what extent do they differ from the material presented or state more than what is obviously apparent from the problem? 1. Low originality 2. Below average originality 3. Average originality 4. Above average originality 5. High originality	How realistic and pragmatic are the consequences and would they occur in the real world? 1. None are realistic 2. A few are realistic 3. Half are realistic 4. Most are realistic 5. All are realistic

It became apparent that further clarification was required to define the difference between obvious and remote responses. Researchers discussed the Guilford and Guilford (1980) definition of obvious responses - those responses directly related to the presented consequence situation. This suggested (although not explicitly defined in the scorer manual) that a remote response is at least a step removed from the situation, not a direct result. For example, "If the world was covered by water" an obvious, direct result would be that everyone would have to learn to swim; while a remote response would be an indirect result, for instance, that swim instructors would cash in (because they get more work because everyone must learn how to swim). This elaboration was not intended to be a deviation from the Guilford and Guilford scoring, but rather a clarification of the original definition of remote responses. Note that a remote response, given this definition, may or may not be a response commonly offered by examinees.

Revisions were made to the remote/obvious categorization scheme provided in the original instrument scorer manual (Guilford & Guilford, 1980). The categorization scheme is a tool to help scorers determine whether a response should be considered remote or obvious. After reviewing current instrument responses, modifications were made to the original categorization scheme to make it relevant to the sample being scored and more user friendly. Upon reviewing the existing categorization scheme, many of the categorized examples were never mentioned in the sample of responses and they seemed to be unlikely responses. For example, it seemed unlikely that the following remote responses would ever be mentioned as consequences to the

problem "*what if everyone lost the ability to read and write.*" Thus, these examples were eliminated from the categorization scheme:

- no formal theater
- more graphite available
- increased use of semaphore.

Many of the examples that were added to the categorization scheme were sports related. Sports were important to the sample, and responses often included some reference to sports activities. Several typical sample responses were added in the categorization scheme.

Several responses were switched to the other category. For instance, one response to "*what if the world was covered with water*," was "there would be increased research on ways of living in the water." Due to the emphasis on indirect versus direct responses, this sample response was switched from remote to obvious. The changes to the rating guide and the scorer training lessened the subjectivity of the ratings, especially since scorers could often find nearly exact matches included in the categorization scheme.

It was also recommended that a two minute time limit be placed on each item. These time constraints were originally proposed by Guilford and Guilford (1980). However, they were not used in the Officer or Civilian administration procedures in which respondents were given a 10 minute time limit to respond to all five questions (or eight minutes to respond to four items in the Civilian study). It was also recommended that an evaluation be conducted to consider the impact of dropping Item 5, the "*what if no one had use of their arms and hands*" problem. The concern was that this situation might be viewed as insensitive to the capabilities of physically impaired individuals.

Alternate Headlines

The *Alternate Headlines* measure presents 10 headlines (see Appendix A). The examinee is instructed to rewrite the headline to retain the original meaning using different words. In the ARI leadership research, each item was rated on Planning, Generation, and Revision. Then, one rating was made for the overall quality and one rating for the overall originality of each examinee's set of 10 headline responses. Presented in Table 2 are the scale definitions that were used when making ratings. The Planning, Generation, and Revision dimensions are components of writing ability identified by Hayes and Flower (1986). The ratings for each of these dimensions were averaged into three overall dimension ratings. Subsequently, the five ratings were combined into two composite measures. The first composite score, Writing Skill, was the average of the Quality, Planning, and Revision ratings. The second composite score, Creative Writing Capacity, was an average of the Originality and Generation ratings.

Since *Alternate Headlines* was originally intended to assess crystallized cognitive skills, HumRRO recommended maintaining separation of the writing skill and creativity concepts.

Thus, three scales were developed to measure Meaning, Presentation, and Creativity. The modified scales are presented in Table 3. Note that the modified Meaning scale taps a similar construct as the original Planning scale. In addition, the modified Presentation scale is similar to the original Generation scale, focusing on the degree to which the headline is well-written.

Table 2. Original Alternate Headlines Scales

Planning:	**Generation**:	**Revision:**
Does the rewrite replicate the key point made in the headline? Is inherent understanding of the meaning of the headline reflected in the rewrite? 1. Rewrite does not reflect the meaning of the original 2. Rewrite somewhat reflects the meaning of the original 3. Rewrite adequately reflects the meaning of the original 4. Rewrite clearly reflects the meaning of the original 5. Rewrite very clearly reflects the meaning of the original	Is the rewrite concise/presented in a headline format? Does the rewrite use words from the original headline? 1. Rewrite is poorly written; uses many words from the original headline. 2. Rewrite is not well written; uses some words from the original headline. 3. Rewrite is well-written/headline format; uses few words from the original headline. 4. Rewrite is concise/effective headline format; uses very few words from the original headline. 5. Rewrite is very concise/highly effective headline format; uses no words from the original headline.	Is the rewrite an effective/better restatement conveying the original meaning of the headline in an equally concise format? 1. Rewrite does no improve on either content or style 2. Rewrite somewhat improves on both content and style 3. Rewrite improves on both content and style 4. Rewrite substantially improves on both content and style 5. Rewrite greatly improves on both content and style
Overall Quality: The extent to which the rewrite reflects coherence and logic with regard to comprehension of the original headline. 1. Low quality 2. Below average quality 3. Average quality 4. Above average quality 5. High quality	**Overall Originality:** The extent to which the rewrite is imaginative and unexpected. The degree of extrapolation from the original statement. 1. Low originality 2. Below average originality 3. Average originality 4. Above average originality 5. High originality	

Meaning and Presentation were combined to form the Writing Skill score, and the Creativity scale represented originality/creativity. It is likely that the instrument is a better measure of creativity than writing skill, per se, since there is little opportunity for examinees to demonstrate the writing processes discussed by Hayes and Flower (1986).

Table 3. Modified Alternate Headlines Scales

Meaning:	**Presentation**:	**Creativity**:
The extent to which the rewritten headline preserves the inherent meaning of the original headline. 0. Blank 1. Rewrite does not reflect meaning of the original 2. Rewrite partially reflects meaning of original 3. Rewrite reflects meaning of original	The extent to which the rewritten headline is well-written (e.g., concise, grammatically correct) 1. Rewrite presentation is poor 2. Rewrite presentation is mediocre 3. Rewrite presentation is average 4. Rewrite presentation is better than average 5. Rewrite presentation is much better than average	The extent to which the rewritten headline is imaginative (e.g., catchy, amusing, emotional), uses words that are different from the original headline, and reflects extrapolation from the original meaning. 1. Rewrite shows no imagination and relies mostly on words from the original 3. Rewrite shows some novelty and does not over-rely on words from the original 5. Rewrite is highly imaginative and most words are different from original

Military Scenarios

Military Scenarios was created specifically for the ARI problem solving leadership research studies. It presents two situations and asks the candidate to answer three questions about each. Eight scales were developed to tap different aspects of problem solving. The eight scales are Short versus Long Term Implications, Attention to Restriction, Nature of Goals - Self, Nature of Goals - Organization, Quality, Objectivity, Number of Alternatives, and Originality, and are described below (Table 4). Ratings from Short versus Long Term Implications, Attention to Restriction, Nature of Goals - Self, and Nature of Goals - Organization were combined to yield a Solution Construction score.

In the Civilian Study, a slightly altered scenario was presented to the civilian examinees. Ratings were made on only four of the scales, Short versus Long Term Implications, Attention to Restriction, Quality, and Originality. An average of the four ratings was used as a Solution Construction measure.

As previously mentioned, it was not clear what part of the leadership model *Military Scenarios* was intended to assess (i.e., one of the 65 KSAPs or a problem solving construct that appears later in the model). Moreover, it appeared that the instrument was not yielding the type of information desired. This was in part because examinees were not really asked to do anything other than repeat back the information provided in the scenarios. This made it difficult to capture any notion of originality or elaboration of the problem because the concepts were vague (making them difficult to operationalize and harder to score) and most responses reviewed did not go beyond the problems as they were presented in the scenario.

A single 7-point scale was developed to measure Problem Construction (Table 5). The Problem Construction score taps the degree to which the examinee effectively frames the overall problem by (a) identifying all the contributing factors that were implied by the scenario description, (b) describing the elements associated with the contributing factors, (c) demonstrating a problem solving perspective by discussing specifically how these factors and factor elements bear on the problem, and (d) identifying and discussing contributing factors that were not specifically identified in the scenario. Although there was some initial concern about

measuring a multidimensional construct with a single scale, it was determined that the 7-point scale would provide an adequate index of the constructs of interest whereas two scales would yield insufficient score variability. Together, the elements of the scale assess the breadth (comprehensiveness) and depth of an examinee's response. Responses to the three questions posed for each scenario were evaluated as a single overall response.

Table 4. Original Military Scenario Scales

Short vs. Long Term Implications: Time span of the goals suggested by the problem/s mentioned. Is the focus on solving mostly the problems at hand (short term goals), or is the focus on issues that might arise as a result of the problem and/or solution or might have caused the problem to begin with (long term implications). 1. Very short term 2. Short term 3. Short and long term 4. Long term 5. Very long term	**Attention to Restriction:** Problem restatement or information requested is attentive to different restrictions in the problem. 1. Does not attend to any restrictions 2. Does not attend to most restrictions 3. Attends to some of the restrictions 4. Attends to many of the restrictions 5. Attends to all the restrictions	**Originality:** Problem restatement or information requested is unique and novel. Extrapolation from the stimulus context, the degree the problem restatement or information goes beyond the rote. 1. Low 2. Below average 3. Average 4. Above average 5. High	**Quality:** Problem restatement or information source is viable for the problem presented. Degree of coherence and logic in problem restatement or information requested. 1. Low quality 2. Below average quality 3. Average quality 4. Above average quality 5. High quality
Nature of Goal - Self Goals focus on one's self, by restating the problem in such a way as to increase the gain or avoid harm to person in question. 1. Not at all self-oriented 2. Slightly self-oriented 3. Self-oriented 4. Fairly self-oriented 5. Very self-oriented	**Nature of Goals - Org** Goals focus on "broader" organizational issues, taking into account multiple constituencies. The problem stated in way to increase broader org. gain. 1. Not at all org-oriented 2. Slightly org-oriented 3. Org-oriented 4. Fairly org-oriented 5. Very org-oriented	**Number of Alternatives** The number of different ways the problem was restated 1. 1-2 2. 3-4 3. 5-6 4. 7-8 5. 9+	**Objectivity:** Problem restatement is attentive to and uses actual facts and key information presented in the problem situation. 1. Not at all objective 2. Slightly objective 3. Objective 4. Very objective 5. Extremely objective

We believe the test could be made appreciably more useful with more substantial revisions to the scenarios and the questions posed to examinees. Given the longitudinal nature of the planned leadership research program, however, it was considered preferable that the test's stimulus materials remain essentially unchanged.

Table 5. Modified Military Scenario Scale

Problem Construction:
The degree to which the overall problem is effectively framed by (a) identifying all the contributing factors that were implied by the scenario description, (b) describing the elements associated with the contributing factors, (c) demonstrating a problem solving perspective by discussing specifically how these factors and factor elements bear on the problem, and (d) identifying and discussing contributing factors that were not specifically identified in the scenario.

1 Response focuses on some of the considerations explicitly cited in the scenario description without detailing elements of those consideration.

2.

3. Response identifies all the key considerations explicitly cited in the scenario description, without discussing the relevance of those considerations or identifies a relevant consideration not cited in the scenario.

4.

5. Response identifies all the key considerations explicitly cited in the scenario description and provides an explicit discussion of how these considerations relate to the solving of the problem or identifies considerations not cited in the scenario.

6.

7. Response identifies all the key considerations explicitly cited in the scenario description, as well as considerations not cited in the scenario and provides an explicit discussion of how these considerations relate to the solving of the problem.

Administrative Modifications

The following modifications are recommended for the administrative procedures for all three instruments.

Test Administration

Only minor changes to the test administration procedures/instruments are recommended. We suggest altering the timing procedures for both the *Alternate Headlines* and *Consequences* tests. In addition, consideration was given to the issue of whether examinees should be told that they would be rated on originality/creativity. Though it was not possible to gather information regarding this issue with the current study (since the responses were from a previous data collection), low variance and validity of the Creativity scores might indicate that examinees should be informed that they will be scored on creativity. Such a change should not be made, however, in the middle of a longitudinal research program.

Scoring Process

Previous research used combined ratings from three scorers to derive scores on the instruments. One purpose of the Demonstration study was to help determine whether three scorers are necessary or whether two scorers would be sufficient for future scoring efforts. To do

this, reliability estimates calculated with three scorers were compared to two-scorer reliability estimates.

Demonstration Plan Procedures

The scale modifications were evaluated using data collected as part of the original Officer study. Specifically, 300 officers in the original study were randomly selected. Their responses to the three instruments and the scores originally assigned in the Officer study were retrieved. Multiple scorers applied the modified scoring systems to the responses. These new (Demonstration) scores were then statistically compared to the original (Archival) scores in terms of descriptive characteristics, reliability, validity, and scorers' reports about ease of use.

The Mumford et al. Officer study was selected as the source for the sample responses because it was the original study that developed and used the instruments of interest, making it the most appropriate baseline for evaluating changes to those instruments or associated scoring procedures (MRI, 1995). Moreover, it was a military sample and data were available on all the instruments. Overall composite and instrument scores from this Archival data set were provided and comparisons were made between the data sets based on score distribution characteristics, reliability estimates reported in the literature, and validity indices. Archival data were not available for individual scorers so interrater reliability estimates could not be re-computed.

The Demonstration plan will be reviewed in the following sections: overview of the demonstration plan design, scorer training, collection of scores and scorer feedback, and data analysis.

Plan Design

The sample of responses used in the Demonstration study is summarized in Table 6. Note that the six officer ranks were combined into three major subsamples of 100 examinees each, for a total of 300 examinees. Note also that each of the examinees in the lower four ranks had responses on all three instruments. *Military Scenarios* data were not available for the O6 examinees and for almost one-half (45%) of the O5 examinees.

Table 6. Sample of Data Responses

Rank	Sample Size	Analysis Sample (Consequences & Alt. Headlines	Analysis Sample (Military Scenarios)
O1 (2nd LT)	50	100 (O1+O2)	100 (O1+O2)
O2 (1st LT)	50		
O3 (CPT)	50	100 (O3+O4)	100 (O3+O4)
O4 (MAJ)	50		
O5 (LTC)	71	100 (O5+O6)	39 (O5+O6)
O6 (COL)	29		
Total	300	300	239

To minimize the scoring burden for individual scorers while maximizing the information provided, examinee responses in the Demonstration sample were divided into eight sets that were distributed among six scorers for scoring. First, the responses were divided into two halves, each with responses from 150 examinees on each of the three instruments (there were fewer responses for *Military Scenarios).* The first half of the analysis sample was divided in a manner that yielded two sets. These were labeled CAM 1 and CAM 2 because they included all three instruments (***Consequences***, ***Alternate Headlines***, and ***Military Scenarios***). The second half of the analysis sample was divided in a manner that yielded six sets. The letter name associated with each set indicates the instrument included in the set (e.g., C1 and C2 included *Consequences* instruments). In summary, the analysis sample was divided into eight sets as illustrated in Table 7 below.

Table 7. Division of Analysis Sample

Response Set (8 total)	Measures Scored	# of Responses
CAM 1	All 3 measures	75 (60 Mil. Scenarios)
CAM 2	All 3 measures	75 (60 Mil. Scenarios)
C1 and C2	Consequences	75 each set
A1 and A2	Alternate Headlines	75 each set
M1 and M2	Military Scenarios	60 each set

Six individuals (two Ph.D. level psychologists and four graduate students) were trained to score responses from the demonstration sample using the modified scoring schemes. The eight analysis sample response sets described above were distributed among the six scorers as shown in the table below. Each scorer rated half the responses (about 150 *Consequences* and *Alternate Headlines* and 120 *Military Scenarios*).

Table 8. Analysis Response Sets

Scorer 1	Scorer 2	Scorer 3	Scorer 4	Scorer 5	Scorer 6
C1	C2	C2	C1	C2	C1
A2	A1	A1	A1	A2	A2
M1	M2	M1	M2	M1	M2
CAM1	CAM1	CAM1	CAM2	CAM2	CAM2

Note that the first half of the analysis sample (Sets CAM1 and CAM2) was rated in a way that minimized variability due to differences in scorers and examinees. That is, a given set of responses was rated by the same three scorers. The second half (Sets C1/C2, A1/A2, and M1/M2) of the sample was rated in a way that maximized variance due to scorers and examinees.

Scorer Training

The six scorers participated in a one-day training session. (Note that two of the six scorers previously served as scorers in earlier validation research.) The training manual is provided in Appendix B. This manual reflects modifications that have been made since training was conducted.

The training included a description of the research program in general, and a discussion of how the ratings would be analyzed and used. It was believed that scorers who have an understanding of how their data will be used will be in a better position to provide valid ratings. Following this introduction to the research study, scorers were trained on each of the instruments. This included a review of the scoring guides and a description of the modified scales for each instrument. For each of the instruments, five responses from the first five officer ranks included in the sample were selected prior to training (too few O6 responses were available to use during training.) These responses were scored prior to training by three individuals familiar with the project and were used for scoring practice during training. The five prescored responses are provided in Appendix C for each of the instruments. In addition to providing the ratings assigned to these instruments, a detailed breakdown is provided for the Consequences responses.

Scorers were trained for approximately two hours on each of the three instruments. For each of the instruments, training followed very similar formats as outlined above. Segments of training that differed for each instrument are described below.

Consequences. During training for *Consequences,* scorers were instructed on how to identify unacceptable and acceptable responses and how to differentiate between obvious and remote responses among the acceptable responses. Scorers were also instructed to treat duplicate responses (i.e., multiple responses covering essentially the same consequence) as one response. Trainers then reviewed several pre-scored responses. Group discussion facilitated the practice scoring process, allowing scorers to hear different perspectives as to what constitutes an obvious and remote response. This discussion resulted in the shifting of several example responses from obvious to remote, and it was decided that if one original response and one remote response were grouped together as duplicates, scorers should classify the response as remote. (Note that the training materials in Appendix B reflect changes made during training.)

Following a discussion of the obvious and remote scales, scorers rated two example responses as a group. Then, two more example responses were rated individually and discussed as a group. Time constraints prevented additional practice scoring. Nonetheless, the scorers expressed confidence in their ability to accurately score the *Consequences* instrument.

Alternate Headlines. Training on *Alternate Headlines* explained how to rate each headline using the three scales. Several questions were raised regarding how to differentiate between the Presentation and Creativity scales. The trainers stressed that the Creativity scale should address the "catchiness" of the response, while the Presentation scale should focus on whether the headline was well-written. Additional questions arose concerning what concepts or ideas were needed to preserve the inherent meaning for the Meaning scale. Trainers and participants went through each of the ten headlines and agreed upon what was required to reflect the fundamental meaning of the headline (these additions are included in the scoring manual provided in Appendix B). Although time constraints again prevented extensive practice scoring, scorers had achieved general consensus when scoring each of the headlines and felt comfortable scoring the instrument.

Military Scenarios. During the training for *Military Scenarios,* trainers discussed the Problem Construction scale. They defined the explicitly-cited factors for each scenario and clarified the difference between *identifying* the factors and *describing* the elements associated with the factors. Furthermore, they explained how to rate responses that identified factors not specifically described in the scenario. One trainee developed a diagram which defined how different types of responses should be scored. This diagram is included in the scoring guide and appears in Appendix B. The discussion of the individually rated responses revealed that all the scorers were confident in their ability to use the Problem Construction scale.

Upon completion of the *Military Scenarios* training, scoring packages were distributed to the scorers. The trainers gave instructions for completing the ratings and addressed final concerns among the scorers.

Collection of Scores and Scorer Feedback

Scorers independently rated responses in the assigned packages. They were allowed to score the instruments at their own pace over a period of four weeks. Scorers were not required to rate the responses in any particular order.

The reaction of scorers to the scoring systems constituted part of the information by which the systems were evaluated. Scorers completed a feedback form after they had completed all of the ratings (see Appendix D). The feedback form contained open-ended questions that examined the scorers' experiences with each of the three instruments in addition to their reaction to the entire scoring experience.

Data Analysis

Data Base Development

As scorers completed their ratings, they faxed them to HumRRO. Data were entered into a spreadsheet format and data cleaning was conducted. Complete descriptive, reliability, and validity analyses were conducted using the Demonstration study ratings. As mentioned, only total scores (not scores assigned by individual scorers) from the Archival data were available for analysis. The data available from the Archival study were the overall dimension scores and scores on external variables. Since Archival data were not available to complete reliability analyses, previously reported reliability estimates were used for comparison purposes.

Data Cleaning

Data were thoroughly examined to ensure that data sets were complete. Out of range data were adjusted by referring back to the original scoring sheets to identify if an error was made in entering the data. In all cases, the out of range data were errors of transcription and were corrected. Data were also checked to ensure that each response had been rated three times. In one instance, a scorer failed to rate a response, resulting in the response having ratings for only two scorers. This response was eliminated from the data set to ensure complete data on all instruments.

Several rules were developed to deal with missing data. If an examinee did not complete the instrument, he or she received a score of zero for each incomplete item on the instrument. On the other hand, if a respondent provided a sarcastic or irrelevant answer, the response was

defined as missing and was not considered in the calculation of the composite scores. This proved to be an infrequent occurrence and consequently did not significantly impact the statistical analyses.

Score Distribution Characteristics

Descriptive statistics were computed for the individual scales and for the composite scores. These statistics provided a basic way to compare ratings from the Demonstration and Archival data. It also provided information about the level of variance in each of the scales.

Score Reliability

Interrater reliability estimates were computed for all basic and composite scores in the demonstration sample. These estimates (intraclass correlations) reflect the average intercorrelation between scores produced by each of the three scorers who rated the response. Exact interrater agreement was examined for the basic scores using a method proposed by Lawlis and Lu (1972).

Internal consistency reliability estimates were computed for the *Alternate Headlines* scores and *Military Scenarios*, though the latter instrument has only two items. *Consequences* is a speeded test with the scores based on the total number of responses developed, so internal consistency is not an appropriate measure of reliability for this measure. Although *Alternate Headlines* is also a timed test, nearly every examinee completed the test and thus, it can be treated as a power test and assessed for degree of internal consistency.

Reliability estimates for the Demonstration data were computed independently for four data sets for each instrument. The four data sets represented the different ways that the responses for each instrument were grouped. For example, the four data sets for *Alternate Headlines* were CAM1, CAM2, A1 and A2 (data sets 1 through 4 respectively). These same four data sets were rated by four different sets of three scorers. Thus, the reliability estimates for each set of responses had to be calculated separately. An overall reliability index for the entire sample (on each instrument) was not feasible since different scorers scored the different data sets.

It was anticipated that these separate analyses would show upper and lower bounds of reliability for the new scoring system. The CAM1 and CAM2 data sets were anticipated to be the higher bound of reliability and the other data sets were projected as the lower bound of reliability. The reliability analyses were also conducted on scores that were calculated using the input of two, rather than three scorers. The goal of these analyses was to determine the extent to which reliability suffered when the number of scorers was decreased.

A special concern with regard to the *Consequences* measure was that the "no use of arms or hands" situation, portrayed by one of the five *Consequences* items, may be insensitive and therefore inappropriate for this type of measure. Accordingly, analyses were conducted to

determine the effects on reliability (and validity) when data from this item were excluded from the composite scores. Analyses were also conducted to examine the impact on the reliability of *Alternate Headlines* when the number of items was reduced from ten to five.

Validity Analyses

While it was not possible to administer additional measures that might have been useful for evaluating the construct validity of the new scores, there were a number of analyses that were performed to evaluate their meaningfulness and sensibility. The major strategy for evaluating the meaningfulness of the various scores was to consider an intercorrelation matrix that included scores (from Demonstration and Archival data) on the three target instruments and scores on other available measures that were expected to be related to the constructs assessed via these instruments. The other scores used for this purpose included rank, verbal reasoning, leader achievement, and a measure of complex problem solving skills.

Demonstration Plan Results

It was difficult to directly compare results from the Demonstration data set to the Archival data because the modifications to the scales resulted in their providing considerably different measures. More specifically, there are fewer new scales than old, so only a portion of the old scales generated in the Archival data set are comparable to the new scales. Moreover, reliability estimates for the Archival scores are only available at the composite score level. Therefore, the composite scores that were conceptually similar across the Demonstration and Archival scales were compared in terms of their statistical performance. Each type of score was compared in terms of score distribution characteristics, reliability estimates, and validity evidence. Following a presentation of the statistical information collected on each instrument, feedback gathered from the scorers is reviewed.

Derivation of Composite Scores

Analyses were conducted on the data to derive composite scores for each of the three instruments. Note that "composite" scores in this discussion refer to scores that were used in subsequent analyses, as opposed to "basic" scores which reflect the scoring scales used by the scorers. Two composites were calculated for the *Consequences* instrument. "Ideational Fluency" was the average of the total obvious scores across all five items. "Creativity" was the average of the total remote scores across all five items.

The three basic scores for *Alternate Headlines* were combined to yield two composite scores. The first composite, "Writing Skill," was computed by summing the average of the meaning items and the average of the presentation items. The second composite, "Creativity," was the average across the creativity items.

The *Military Scenarios* instrument contained one composite score, "Problem Construction," which was the average score across the two scenarios.

Score Distribution Characteristics

Table 9 shows the means and standard deviations of comparable Archival and Demonstration data set scores. Note that the sample size of the Demonstration and Archival data sets was about 300, except for *Military Scenarios* which was 239. As mentioned previously, there were fewer higher ranking officers who completed the *Military Scenario* instrument.

Table 9. Instrument Descriptive Statistics

Variable Score	Sample Size	Mean	Std. Dev.	Range of Reliability (3-scorers)	Range of Reliability (2-scorers)
Demonstration data sets					
Alt. Headlines: Creativity	300	2.69	.52	.89 - .92	.80 - .91
Alt. Headlines: Writing Skill	300	5.05	.46	.78 - .90	.64 - .90
Conseq: Ideational Fluency	301	1.74	.64	.82 - .92	.67 - .92
Conseq: Creativity	301	1.91	.76	.86 - .94	.77 - .93
Mil. Scen: Problem Construction	239	3.09	.79	.85 - .90	.72 - .91
Archival data set					
Alt. Headlines: Creativity	300	2.52	.40	.82, .57	
Alt. Headlines: Writing Skill	300	2.28	.35	.68	
Consequences: Originality	301	2.79	.46	.82	
Consequences: Quality	301	2.64	.47	NA[1]	
Mil. Scen: Solution Construction	239	3.01	.53	.67, .79	

The Writing Skill score in the Demonstration data set was the sum of two scales, Presentation and Meaning. The mean for Writing Skill was higher than the other variables, however, the standard deviation remains the lowest of all the measures in the Demonstration data set. Considering the standard deviation descriptive information for Presentation and Meaning (since it is a 3-point scale) in Appendix E, it appears that the standard deviation for Writing Skill may be a function of the relatively low standard deviation of the Presentation scale.

[1]Reliability estimates for the *Consequences* "Quality" scale could not be located.

Score Reliability

Interrater Reliability

Table 9 also provides reliability estimates for the Demonstration and Archival data set scores. The individual data sets are described in terms of a range of reliability scores[2]. For each instrument, Demonstration data sets 1 and 2 represent ratings of the CAM1 and CAM2 data sets. Recall that these data sets were expected to have reduced variability associated with scorers and examinees. Data sets 3 and 4, for each instrument, represent ratings for the C1/C2, A1/A2, and M1/M2 data sets, which were expected to have greater variability associated with scorers and examinees. Complete results for each data set are presented in Appendix E. The individual data set scores did not show the expected differences in variability.

With regard to the scores in the Archival data set and the Demonstration data sets that were conceptually comparable (e.g., the old and new Creativity scores from *Alternate Headlines*), the reliability estimates for the modified scores (Demonstration data set) were consistently stronger, showing an improvement over the Archival scores. The whole range of reliability estimates for every Demonstration data set score was higher than the associated scores' reliability estimates from the Archival data. Interrater reliability estimates for all data sets are provided in Appendix E.

The largest increase in reliability (from Archival data to Demonstration data) occurred for *Military Scenarios* and *Alternate Headlines* (Writing Skill) scores, even though Writing Skill includes the lowest reliability estimates relative to the other Demonstration scale composites. As discussed above, the standard deviation for the Writing Skill score was surprisingly low. It is likely that the attenuated standard deviation for Writing Skill impacted its reliability estimates as well.

Also included in Table 9 are reliability estimates for the Demonstration data set using two scorers rather than three. The two-rater reliability estimates were calculated using all possible combinations of two scorers, and the range of scores across all four data sets are presented. While the range of reliability estimates for the two-scorers overlapped to a good degree with the three-rater estimates, the range of scores for two-scorers included lower reliability estimates and was much more variable. This indicated that the specific combination of two-scorers greatly impacted the resulting reliability estimate.

Interrater reliability analyses were also conducted to determine the impact of dropping Item 5, "*What would be the results if no one could use their arms or hands*?" on the reliability of the *Consequences* instrument. Table 10 shows the reliability estimates for both composite scores

---

[2]Reliability estimates reported for the Archival data set were drawn from Tremble et al. (1997) and the *Cognitive and Temperament Predictors of Executive Ability: Principles for Developing Leadership Capacity*, a presentation by MRI and George Mason Center for Behavioral and Cognitive Studies (1995).

for the original length *Consequence* measure and the shortened measure, with Item 5 eliminated. The impact of eliminating Item 5 was greater on the Ideational Fluency scale than on the Creativity scale. The decrease in reliability on the Ideational Fluency scale was not so extreme, however, as to eliminate the shortened scale from consideration. Furthermore, for the Creativity Scale, the reliability remained the same in data set 4 and actually increased in data set 1 when Item 5 was dropped.

Table 10. Reliability of Shortened *Consequences* Measure

	Ideational Fluency - Obvious Scale		Creativity - Remote Scale	
Data set number	Original length	Elim. #5	Original length	Elim. #5
1	.83	.78	.86	.88
2	.87	.84	.92	.91
3	.82	.77	.92	.91
4	.92	.89	.94	.94

Table 11 shows the impact on reliability estimates for *Alternate Headlines* when the number of items in the scale was reduced from 10 to 5 items. The first 5 items were retained to comprise the shortened scale. The change in reliability for the Writing Skill scale ranged from decreasing by .14 to increasing by .01. The reliability for the Creativity scale remained fairly constant for the shortened scale for data sets 1, 2, and 3; however, the reliability dropped by .09 for data set 4.

Table 11. Reliability of Shortened *Alternate Headlines* Measure

	Writing Skill		Creativity	
Data set number	Original length	Elim. 5 items	Original length	Elim. 5 items
1	.78	.64	.89	.89
2	.81	.82	.90	.87
3	.90	.85	.92	.91
4	.87	.86	.91	.82

Interrater Agreement

Lawlis and Lu's chi-square analysis (1972) was used to calculate interrater agreement. This technique allows the researcher to compute interrater agreement according to a predetermined criterion. The level of agreement is computed by specifying the number of items

being rated, the number of observed agreements and disagreements, the number of raters, the number of points on the scale, and the probability of *k* judges achieving agreement by chance. Significant results provide evidence that any observed agreement is not due to chance.

The interrater agreement indices for *Alternate Headlines* and *Military Scenarios* are presented in Appendix F. Indices of agreement for *Consequences* were not possible using the Lawlis and Lu methodology since this index estimates the level of agreement when scorers make judgments based on a scoring scale and *Consequences* scores were based on frequency counts. The index of agreement was significant for nearly all ratings for 3-scorers (and 2-scorers) for *Alternate Headlines* and *Military Scenarios*. Because almost all scales and items showed significant levels of agreement between scorers, there was no evidence that one or more of the scales was difficult to use or introduced greater subjectivity into the ratings.

Internal Consistency Reliability

Appendix G shows coefficient alpha estimates for each *Alternate Headlines* scale, overall and with each item removed from the scale. The overall coefficient alphas range (across the four Demonstration data sets) from .68 to .81 for the Meaning scale, .60 to .67 for the Presentation scale, and .84 to .89 for the Creativity scale. Again, it is only the Presentation scale that has lower than desired reliability. However, estimates calculated excluding individual items did not suggest that improvements could be made by eliminating outlier items.

Internal consistency for the *Military Scenarios* measure was estimated by the correlation between the two scenarios included in the instrument. This correlation was .45, indicating a moderate relationship between the two items. Given that this internal consistency estimate is based on only two items, .45 does not seem to be unreasonable.

Score Validity

Table 12 is a correlation matrix that describes the relationships among comparable scores within and between the Demonstration and Archival data sets and a set of external variables. In addition to scores on *Alternate Headlines*, *Consequences*, and *Military Scenarios*, the original Officer study (the source responses scored in both the Demonstration and Archival data sets) included data on several other variables. Among these were several variables that theoretically should correlate with scores on the target measures. These external variables were rank, verbal reasoning, leader achievement, and selection of solution components. Rank and leader achievement are indicators of leader effectiveness. Verbal reasoning, measured by a subtest of the Employee Aptitude Survey (EAS), is a measure of cognitive capabilities (Tremble et al., 1997). Selection of solution components is a measure of complex problem solving skills developed by Mumford et al. (1993).

Table 12: Variable Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
Demonstration													
1 AH-Creat													
2 AH-Wri Ski	-23**												
3 Csq-Idea Fl	04	02											
4 Csq-Creat	20**	01	15*										
5 MS-ProCon	16*	16*	09	31**									
Archival													
6 AH-Create	58**	18**	05	21**	21**								
7 AH-Wr Sk	-16**	70**	-004	11	21**	48**							
8 Csq-Orig	14	11*	29**	61**	41**	40**	27**						
9 Csq-Qual	12*	16**	31**	58**	44**	33**	27**	88**					
10 MS-SolCon	07	19**	12	25**	82**	26**	28**	51**	51**				
External Vars													
11 Verbal Rea	27**	25**	12	18**	11	31**	23**	25**	19**	08			
12 Leader Ach	-02	14*	04	15**	34**	20**	27**	42**	44**	42**	07		
13 Rank	00	22**	10	16**	38**	26**	33**	53**	57**	47**	10*	78**	
14 SelSol	02	-01	-04	21**	20**	10	03	25**	24**	19**	06	27**	21

Note. * = p<.05; ** p<.01 (two-tailed).

Consequences

The *Consequences* Ideational Fluency scale was moderately correlated (r=.31) with the Archival *Consequences* Quality score. It was not related to verbal reasoning or any of the other external variables. The *Consequences* Creativity scale was more highly correlated (r=.61) with its most conceptually comparable score in the Archival data set, *Consequences* Originality. Interestingly, the Demonstration Creativity score was also more highly correlated with the Archival Quality score (r=.58) than the more conceptually equivalent Ideational Fluency score. The Creativity score was also more highly correlated with the external variables than was the Ideational Fluency score. It was particularly highly correlated with the problem solving measure. Both of the Demonstration *Consequences* scores, however, showed lower correlations with the external variables than the Archival scores.

To further examine the impact of removing the "arms and hands" item from the *Consequences* measure, correlations were recomputed using *Consequences* scores that were based only on the other four items. For the most part, the correlations remained essentially the same, deviating by only one- or two-hundredths of a point. In a few cases, the correlations actually increased with the removal of the item.

Alternate Headlines

The Demonstration *Alternate Headlines* Creativity score was reasonably correlated with the Archival *Alternate Headlines* Originality score (r=.58). The Demonstration *Alternate Headlines* Writing Skill score, however, was even more strongly correlated with the Archival Writing Skill score (r=.70). Both scores were correlated with verbal reasoning (r=.27 and .25, respectively). Indeed, the Writing Skill score was significantly correlated with all the external variables except for Selection of Solution Components. However, Creativity was not correlated with any external variables. Except for the correlation with verbal reasoning, all the correlations between the *Alternate Headlines* scores and the external variables are higher for the Archival scores than the Demonstration scores.

Military Scenarios

The Problem Construction scale for the *Military Scenarios* instrument (Demonstration data set) was highly correlated (r=.82) with the Archival *Military Scenarios* composite Solution Construction score. Both scores significantly correlated to leader achievement, rank, and Selection Solution, with the former two correlations somewhat higher for the Archival score compared to the Demonstration score. Neither the Demonstration or the Archival *Military Scenarios* score was correlated with verbal reasoning.

Creativity Measures

Creativity is a construct that both *Alternate Headlines* and *Consequences* were intended to assess. Within the Demonstration data set, the *Alternate Headlines* Creativity scale correlated significantly, though not strongly, with the *Consequences* Creativity scale (r=.20). There was a somewhat higher correlation between their counterparts in the Archival data set (r=.40).

The Influence of Rank

The correlations between the external variables and the Archival scales in many cases were stronger than the external variables with the Demonstration scales. This might suggest that, despite improved reliability, the Demonstration scales have lower predictive validity than the Archival scales. One likely contributor to the differences between the correlations was a difference in the Archival and Demonstration study methodologies. Aside from the modifications to the instrument scoring schemes in the Demonstration study, the scoring processes used in each study also differed. Specifically, scorers in the Archival study were not

blind to examinee rank, whereas scorers in the Demonstration study were. This may have confounded the Archival scale ratings, thus, inflating the correlations between the instrument scores, rank, and leader achievement (which was highly correlated with rank).

HumRRO examined this hypothesis by re-computing correlations with the influence of rank partialed out. Table 13 reprints the affected portion of the Table 12 correlation matrix including the partial correlations instead of the simple correlations. Note that the correlations with verbal reasoning are essentially unaffected, presumably because the EAS subtest was objectively scored. Most correlations with the other external variables, both with the Demonstration and the Archival data sets, were reduced. Notably, *no* scores were significantly correlated with leader achievement when rank was partialed out, and any predictive advantage the Archival *Military Scenarios* score had over the new score was eliminated.

Table 13: Correlations with Rank Partialed Out

Variable	1	2	3	4	5	6	7	8	9	10	11	12
11 VerbRea	27**	26**	11	17**	12	31**	24**	26**	19**	09		
12 LdrAch	-03	-04	-05	05	08	00	02	-02	02	12	-05	
14 SelSol	02	-05	-06	19**	12	06	-03	17**	15**	10	-02	18

Note. * = p<.05; ** p<.01 (two-tailed).

Scorer Feedback

The individuals who served as scorers during the Demonstration study completed feedback forms describing their experiences using the modified instruments (see Appendix D). They were asked to comment on each of the instruments individually and to provide feedback regarding the training workshop and the scoring process. With regard to general comments, there was a suggestion to provide more thorough training on each individual scale. For instance there was a suggestion to provide a more explicit, detailed explanation and review of what all the anchors on each of the scales mean. This was done to a great extent, but additional time may be integrated in the future to elaborate this section of training. The feedback specific to each of the individual instruments is provided below.

Consequences. The *Consequences* measure was considered by most scorers to be the easiest instrument to score. The scoring guide, which included a scoring key with examples of how responses should be rated (i.e., remote versus obvious), was viewed as a great help to the scorers. Scorers did recommend including additional discussion during training about what is considered an irrelevant response, how to treat responses that are abstract rather than specific, and what is considered duplicate responding (i.e., a series of similar responses).

Alternate Headlines. The Meaning scale was considered the most objective and easiest scale to use of the three *Alternate Headline* scales. The Presentation scale, on the other hand, was the most problematic. The Presentation scale was defined as "how concise and grammatically correct is the response?" Scorers reported that the two dimensions, being concise and being grammatically correct, were often at odds with each other. To be grammatically correct and maintain the original headline's meaning often prevented a response from being concise. Thus, the Presentation scale was difficult to rate. In addition, scorers were told to use a rating of 3 as the baseline for responses being grammatically correct. With a rating of 3 indicating an acceptable response in terms of grammar, however, scorers had difficulty conceptualizing what an excellent response would look like. Therefore, scorers indicated that they did not use the whole scale.

Military Scenarios. Most scorers indicated that the multidimensionality of the Problem Construction scale made it difficult to rate the responses. They liked the instructional diagram developed during training because it was helpful in integrating the information to make a rating. Several scorers also noted that the *Military Scenario* questions are not targeted to elicit the information that is rated by the Problem Construction scale (a problem that was even more severe with the original scoring scales). There was also a suggestion to provide examples of responses that illustrated different ratings, particularly in terms of external considerations related to each scenario.

Recommendations for Modifying Systems

The objective of the current project was to improve the reliability of the *Alternate Headlines, Consequences,* and *Military Scenarios* instruments. Given this objective, scoring modifications for each of the instruments were implemented. In order to evaluate the results using the scoring system modifications, results of a Demonstration study were compared to Archival scores based on the original scoring systems used in ARI's leadership research. This section discusses the results of the Demonstration study and provides final recommendations for modifying each instrument.

General

With the exception of the Presentation scale for scoring *Alternate Headlines*, the Demonstration study supports adoption of the modified scoring systems for all three instruments. The new scores exhibit higher levels of reliability than the old scores. Moreover, the modified scoring systems are considerably simpler than the scoring used in prior Army leadership research, both in terms of the number of dimensions to be rated and the conceptual distinctions among dimensions to be rated. The improvement in the conceptual distinctions among dimensions was gauged based on a comparison of comments made by Archival and Demonstration scorers.

With regard to training, scorers were generally interested in having more time to review the scoring scales and example responses. For example, one suggestion for future training on *Consequences* would be to provide more examples of irrelevant and duplicate responses. Therefore, the training schedule should allow for at least three hours per instrument, compared to the approximately two hours per instrument provided in the Demonstration study. It may also be useful to select additional sample responses (other than those presented in Appendix C) to use in training.

One of the issues considered was whether to tell future examinees that their responses will be rated on creativity. It was suggested that this could favorably affect both score validity and score distribution characteristics. As with the possibility of revising the instruments themselves, however, substantively changing administration conditions was not desired given the planned longitudinal nature of the research program. Moreover, the creativity measures showed reasonable descriptive characteristics during the Demonstration study and the predicted relationships with other measures of creativity and complex problem solving skills were found. Thus, it is recommended that future administration procedures be changed only in the very minor ways suggested earlier in this report (e.g., separately timing the *Consequences* items).

Across all three instruments, ratings that were computed based on two scorers instead of three showed considerable variability in interrater reliability depending on the scorers that were paired. Since the range of two-rater reliability estimates dipped fairly low in some cases, it is recommended that future scoring activities continue to use three scorers.

Consequences

The modified *Consequences* scoring system, which was essentially the same as that originally developed by Guilford and Guilford (1980), exhibited higher reliability estimates than the scoring system developed by MRI (1995). It is also much easier and less time-consuming to score than the system used in the earlier leadership research because it involves counting responses, with minimal judgment about those responses required, whereas the earlier system required subjective ratings on up to eight dimensions. This simpler system, moreover, retains the ability to measure creativity as desired for testing the problem solving leadership model.

Due to sensitivity concerns about Item 5 (i.e., the no use of arms or hands item), the impact of eliminating that item on reliability estimates was examined. The Ideational Fluency scale experienced a slight decrease in reliability across data sets, while the Creativity scale either remained fairly stable or increased across the data sets. There also seemed to be little if any impact on the predictive validity of the shortened measure, as demonstrated by its correlations with external variables. Thus, we recommend that this item be dropped given that there is no great statistical advantage in retaining the item. Dropping the item will avoid the potential of offending examinees and will reduce the scoring burden for this instrument.

Alternate Headlines

The original scoring scales for this instrument were designed to tap writing skill components (i.e., planning, generation, and revision) that would be demonstrated in an essay, but are not likely to manifest themselves in brief headlines. The modified scoring system is an improvement over the original because it is more reliable and tailored more effectively to the instrument itself. The first scale asks whether the examinee has preserved the meaning of the headline (as instructed in the directions), and the other two scales examine grammatical correctness and creativity. The second scale, however, still proved problematic, both in terms of psychometric characteristics (reliability and validity) as well as input from scorers.

Scorers indicated that they had difficulty using the 5-point Presentation scale. Most scorers indicated that they did not use the whole scale range, which resulted in the low standard deviation for the scale. This also impacted the reliability estimates for Writing Skill. Table 14 presents a new version of this scale that is intended to reduce the identified problems.

Table 14. Newly Modified Presentation Scale

Presentation:
The extent to which the rewritten headline is grammatically correct (e.g., agreement of subject and verb) and is written in headline format. Effectively formatted headlines avoid past tense and do not reveal details or extraneous information about the article (e.g., specific names).

1. Headline is not grammatically correct AND (uses past tense OR provides extraneous information that makes the headline too long).
2. Headline is not grammatically correct OR uses past tense AND provides extraneous information that makes the headline too long.
3. Headline is grammatically correct but is written in past tense AND contains extraneous information
4. Headline is grammatically correct but is written in past tense OR contains extraneous information
5. Headline is grammatically correct, avoids past tense, AND contains only essential information

As illustrated by the new scale, it is recommended that the scale continue to address grammatical issues while also considering the format of the headline. That is, a good headline is concise in addition to avoiding past tense, since a strong headline should reflect the news in the present or future. Although scorers felt that these two dimensions were often at odds with each other, providing concrete scale anchors as shown here should reduce the confusion created by the previous scale. Scorers also indicated that they had difficulty conceptualizing what an excellent response would look like. The new scale defines the combination of grammatical and headline format characteristics required for each possible rating, much like the multidimensional Problem Construction scale used for *Military Scenarios*.

A shortened *Alternate Headlines* instrument was also examined. The results using only the first five items provide some optimism that the Creativity scale reliability might hold; however, the impact on the Writing Skill scale is more significant. Although it is recommended that all ten items continue to be scored, the impact of scoring fewer items should be revisited once data on the newly modified Presentation scale are available.

Military Scenarios

Although the Problem Construction scale exhibited higher reliability than the comparable composite variable from the Archival data set, scorers continued to be concerned about the multidimensionality of the scale. They found it somewhat difficult to conceptualize both dimensions on the same scale. There was also some concern that the rating of "3" had many different meanings.

Despite concerns about the scale's multidimensionality, it is recommended that the single scale approach be maintained as long as the instrument does not change. Problem Construction, as a construct, is multidimensional, and includes consideration of factors explicitly addressed in the scenario, factors beyond the information provided, and some integration and construction of how the factors relate and impact on one another and the situation. The questions following each scenario fail to prompt examinees to elaborate on the information they consider relevant in the presented situation. Thus, the responses provided often lack variability on the individual dimensions, particularly in terms of the elaboration of important factors that influence the situation. Because of this lack of response variance, attempting to extricate two or more factors from the responses, although desirable in concept, is not feasible. The strategy of using one scale, incorporating all the problem construction dimensions, is recommended as the best method for capturing the response variance that is there to be captured.

Because of the scale's complexity, however, future training should continue to provide scorers with the diagram that helps them visualize the prospective options. Scorers should be reassured that several different types of responses can get the same rating and that this is not a problem since the scale was designed to be multidimensional. That is, a response lacking breadth *or* depth will receive a low score. Further, trainers should emphasize that the instrument does not require examinees to elaborate or discuss factors regarding the presented problem. Without changing the instrument itself, some candidates will continue to receive low scores because they are responding specifically to the questions asked, rather than elaborating on the issues as we hope they will. Presumably, this will reduce the validity of the scores, but the problem is not fully correctable with the scoring system alone.

A final comment about the *Military Scenarios* instrument. Despite the difficulty that scorers have had with this instrument, using both the Archival and Demonstration scales, it has consistently shown the highest correlations with external variables. As pointed out above, the major problem inherent in the instrument is that it was developed to measure something more complex than what it actually asks respondents to demonstrate. Examination of the three

questions posed to examinees shows that they are just asked to repeat back what they have read in the scenario. Examinees who choose to go beyond what they have been asked and elaborate upon the scenario presented, may be showing initiative. This initiative, in turn, may be what is related to important external variables. This possibility could be explored in research that compares the current version of the instrument with a version that poses questions that are more in line with the goals of the instrument.

References

Guilford, J.P. (1966). Intelligence: 1965 Model. *American Psychologist, 21*, 20-26.

Guilford, J.P., & Guilford, J.S. (1980). *Consequences sampler set: Manual of instructions and interpretations.* Palo Alto, CA: Mind Garden.

Guilford, J.P., & Hoepfner, R. (1966). *Structure of intellect factors and their tests.* Los Angeles, CA: University of Southern California, Psychological Laboratory.

Hayes, J.R,. & Flower, L.S. (1986). Writing research and the writer. *American Psychologist, 41*, 1106-1113.

Lawlis, G. F., & Lu, E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin, 78*, 17-20.

Management Research Institute (1995). *Cognitive and temperament predictor of executive ability: Principles for developing leadership capacity.* (Briefing to the U.S. Army Research Institute) Bethesda, MD: Management Research Institute.

Mumford, M.D., Zacarro, S.J., Harding, F. D., Fleishman, E.A., & Reiter-Palmon, R. (1993). *Cognitive and temperament predictors of executive ability: Principles for developing leadership capacity* (ARI Technical Report 977). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A267 589)

Tremble, T.R., Jr., Kane, T.D., & Stewart, S.R. (1997). *A note on organizational leadership as problem solving* (ARI Research Note 97-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Appendix A - Predictor Instruments

Consequences

In this test you will be presented five difference questions. For each question, you are to generate as many responses as you can in the time allowed for each question.

Below is an example of the questions in this test. As you can see, the example has four sample responses to get you started. There are also spaces for you to write in your own responses. The five test questions are like the example below.

You will have two minutes to respond to each question. For each question, please start and stop as instructed.

EXAMPLE:

What would be the results if people no longer needed or wanted sleep?

SAMPLE RESPONSES:

Get more work done.
Alarm clocks not necessary.
No need for lullaby song books.
Sleeping pills no longer used.

YOUR RESPONSES:

1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

STOP
PLEASE DO NOT TURN THE PAGE UNTIL INSTRUCTED.

QUESTION 1.

What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water, except for a few of the highest mountain peaks?

Sample Responses:

a. Everyone will move to the mountain peaks
b. Increased sale of boats
c. Business failure
d. Panic

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

10 _____

11 _____

12 _____

13 _____

14 _____

15 _____

16 _____

STOP

PLEASE DO NOT TURN THE PAGE UNTIL INSTRUCTED.

QUESTION 2.
What would be the results if suddenly everyone lost the ability to read and write?

Sample Responses

a. No newspapers or magazines
b. No libraries
c. No mail or letters
d. T.V. sales increase

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

10 _____

11 _____

12 _____

13 _____

14 _____

15 _____

16 _____

STOP
PLEASE DO NOT TURN THE PAGE UNTIL INSTRUCTED.

QUESTION 3.

What would be the results if human life continued on earth without death?

Sample Responses

a. Overpopulation
b. More old people
c. Housing shortage
d. No more funerals

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

10 _____

11 _____

12 _____

13 _____

14 _____

15 _____

16 _____

STOP
PLEASE DO NOT TURN THE PAGE UNTIL INSTRUCTED.

QUESTION 4.

What would be the results if the force of gravity was suddenly cut in half?

Sample Responses

a. Jump higher
b. More accidents
c. Less effort to work
d. Easier to lift things

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

10 _____

11 _____

12 _____

13 _____

14 _____

15 _____

16 _____

STOP
PLEASE DO NOT TURN THE PAGE UNTIL INSTRUCTED.

QUESTION 5.

What would be the results if no one could use their arms or hands?

Sample Responses

a. Learn to use feet more
b. No need for gloves
c. Clothing would be changed
d. Couldn't drive cars

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

10 _____

11 _____

12 _____

13 _____

14 _____

15 _____

16 _____

STOP
PLEASE DO NOT TURN THE PAGE UNTIL INSTRUCTED.

SSN _____

ALTERNATIVE HEADLINES

This test provides you with 10 headlines from major newspaper clippings. Please rewrite each headline to say the same thing using different words. Remember, your rewrites should be in the headlines format. Please wait to start until directed to start.

Sample Headline: FAMILY PET STALKS GUNMAN

Sample Rewrite: FIDO SAVES THE DAY: ROBBERY THWARTED

You will be allowed 10 minutes for the whole test, which means you have about one minute for each "rewrite." To help you keep track of the time, the test administrator will count off time remaining in two-minute intervals.

1. Headline: MAN DROWNS IN VAIN EFFORT TO SAVE FIANCEE

   Rewrite:

2. Headline: PLANES COLLIDE OVER OCEAN, KILLING THREE

   Rewrite:

3. Headline: GRANDMOTHER SURVIVES ANOTHER DESERT JET CRASH

   Rewrite:

4. Headline: CAR THIEVES ARE DUPING SELLERS BY 'TESTING' AND NOT RETURNING

   Rewrite:

5. Headline: MORE COMMUTERS ARE TRAVELING ALONE

Rewrite:

6. Headline: STATE INMATES KEEP RURAL HIGHWAYS CLEAN

Rewrite:

7. Headline: AUTO BEING CHASED BY POLICE HITS VAN

Rewrite:

8. Headline: STUDY FINDS LOW-INCOME YOUNGSTERS GET INADEQUATE ATTENTION

Rewrite:

9. Headline: SUMMER PROGRAMS MAKE LEARNING REWARDING

Rewrite:

10. Headline: RESEARCH TEAM FINDS VACCINE THAT COULD SAVE MILLIONS

Rewrite:

SSN _____

Military Scenarios

Problem One

You are a corps commander whose forces are tired and shorthanded as a result of several weeks of rugged fighting. A significant number of your troops are inexperienced replacements who have recently been assigned to their units. Although you have been given some intelligence indicating that the enemy may be massing troops on your front, you do not consider it a serious threat due to the overall state of the war. You have assured your superiors that you have adequate strength to defend your sector, and have resisted suggestions that our forces be folded into a neighboring command with fresher troops. However, it soon becomes apparent that the enemy is mounting a major effort.

Questions:

1. If you were placed in this situation, what would be the most important problem for you to address?

2. What key pieces of information would you need to solve the problem?

3. Are there other problems you would have to consider?

Problem Two

Upon retirement from the military service, you have been appointed head of a large government agency in your home state. Your appointment came about in part from your success and achievements during a distinguished military career, which was well publicized in the local press. You view this as being the first step in what you hope will be a long career of appointed or elected government service. However, for this to happen, you must demonstrate that you can be successful in a politically charged civilian environment. In recent years, your agency has not been as effective as it might have been in dealing with problems which have now become quite serious and require decisive action. Observers of governmental operations within the state have felt that your predecessor had little influence with the Governor in developing policy in your agency's area of responsibility. As a result of this, the productivity and morale of the employees within your department is very low. In addition, budgetary constraints resulting from lower state revenues will affect your efforts to improve your agency's contribution to the state's welfare.

Questions:

1. If you were placed in this situation, what would be the most important problem for you to address?

2. What key pieces of information would you need to solve the problem?

3. Are there other problems you would have to consider?

Appendix B - Scorer Training Manual

COGNITIVE MEASURE SCORING DEMONSTRATION PLAN
SCORER TRAINING

February 19, 1997

8:00 Review purpose of project & demonstration plan methodology

9:00 Scorer training for *Consequences*
- Review scoring system using 5 pre-scored sample responses
- Rate 2 sample responses as a group
- Rate 2 sample responses individually, then discuss as a group; repeat if necessary

11:00 Scorer training for *Alternative Headlines*
- Review scoring system using 3 pre-scored sample responses
- Rate 2 sample responses as a group

12:00 Lunch

1:00 Scorer training for *Alternative Headlines* (Continued)
- Rate 2 sample responses individually, then discuss as a group; repeat if necessary

2:00 Scorer training for *Military Scenarios*
- Review scoring system using 5 pre-scored sample responses
- Rate 5 sample responses as a group
- Rate 5 sample responses individually, then discuss as a group; repeat if necessary

4:00 Distribute rating packages (examinee responses, score sheets, feedback forms); review procedures and timelines for returning completed materials

Consequences

Consequences Scoring Process

Each of five different questions will be rated on two dimensions: originality and ideational fluency. Two scores will be generated for analysis.

Originality: The extent to which the consequences are novel and imaginative and the extent to which they differ from the material presented stating more than is obvious in the problem. An original or remote response is indirectly related to the stated problem. The original response is not an obvious or direct result, but requires an additional step in the thought process to arrive at the indirect, secondary consequence of the problem.

Total number of responses that are considered "remote." *(Rate the item 0 if the item is blank)*

CR Cannot rate the item because responses cannot be taken seriously or is an unacceptable response.

Ideational fluency: The number of obvious, acceptable (relevant, non-duplicate) responses. An obvious response is a direct result of the problem.

Total number of responses that are considered "obvious." *(Rate the item 0 if the item is blank)*

CR Cannot rate the item because responses cannot be taken seriously or is an unacceptable response.

Consequences Scoring Aid

The following scoring guide outlines responses that can be considered remote and obvious.

CONSEQUENCES SCORING GUIDE

Two scores are computed based on responses to this exercise, ideational fluency and originality. Ideational fluency is a count of the total number of responses that are acceptable and fall into the obvious category. Originality is a count of the total number of responses that are acceptable and fall into the remote category.

Unacceptable Responses

Unacceptable responses, defined as irrelevant responses and duplicate responses, are not included in the scoring of the test.

Irrelevant responses. Irrelevant responses are those that are not germane to the question. A sarcastic response, or a response which indicates that the respondent is not seriously responding to the question would also be considered irrelevant. Caution should be taken in labeling a response as irrelevant since at first glance it may appear irrelevant but with further consideration may turn out to be pertinent, and perhaps a very remote response.

Duplicate responses. A duplicate response may be a repeat of one of the examples, or a rewording of a previously presented idea. When examinees identify a category of responses, (e.g., kinds of schools, or kinds of transportation) and several responses belong to the same category, the set of responses should be counted as one response. For instance, the following would be redundant responses for "What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water..."

More masks sold
More oxygen tanks sold
More snorkels sold
More underwater cameras sold
More swimming attire sold.

Each of these responses is related to the theme of selling swimming/diving gear and thus are considered redundant and counted as one response.

Acceptable Responses

If responses are not irrelevant or duplicate, as described above, they are considered acceptable responses and are included in one of the two test scores.

Ideational Fluency. There is an overall **ideational fluency** score which is the total number of obvious-acceptable responses. An obvious response is one that indicates a direct, immediate result of the changes suggested in the question, or refers to a commonly associated function. This response may be quite typical and is an *obvious* result of the situation presented.

Originality. An **originality** score is based on the total number of remote-acceptable responses. Remote responses describe results that are not an obvious, direct result of the situation, but are more distant temporally or geographically, or indicate a specific substitute, a new system, or some other fairly specific way of adjusting to the changed situation.

Scoring guide. Sample obvious and remote responses are provided for each question. The responses are provided in general categories. The scorer is encouraged to compare the responses to the sample to help determine whether responses are obvious or remote. If a response does not appear on the list, the scorer should attempt to relate the given response to those provided. Responses should not be rejected just because they do not appear on the list.

Item 1 - World covered with water	
Remote	**Obvious**
Solutions	
cities built on stilts	people live on boats
underwater refuges built	food is stored
build artificial islands	more boats built
migrate to other planets	everyone learns to swim
space program really takes off	increased research on ways of living in water
Physical Consequences	
uniform world temperature	loss of homes
division of nations/states disappear	people drown and otherwise die
nautical maps at a premium	loss of pets
new forms of sea life emerge	
Social Consequences	
decline in morals	businesses/schools shut down
suicide rate increases	fishing gear sales increase
religious revival	panic
no one believes in Bible	bathing suit and scuba gear sales increase
many people cash in on insurance	killing each other to reach peaks
sea sickness medication sales increase	mountain real estate increases value
bath tubs no longer needed	increased water sports
people work together to be safe on mountain peaks	increase water movie rentals

Item 2 - Loss of reading and writing ability	
Remote	**Obvious**
Solutions regarding communication	
signs identified by shape	more radios
increase in picture book sales	more movies and more TV
telepathic communication developed	increased use of telephones
interactive e-mail systems	more oral communication
Consequences in literacy area	
no money	no more books
no accurate history	no plays or scripts
more automobile accidents	decrease in communication
no more illiteracy	no tests
problems with public transportation	no bedtime books
directions difficult to follow	no need for pencils or pens
decline in businesses associated with writing instruments	
Social Changes	
folklore becomes more popular	no authors or poets
more leisure time	no school
intelligence drops	lack of classic literature
adults and children are equal	no more newspaper/magazines
crime increases	no more billboards
fewer jobs	
everything paid for in cash	
race riots	
decrease in demand on lumber	
Other Changes	
increase in noise pollution	
computer sales drop	

Question 3: No Death	
Remote	**Obvious**
Solutions to Overpopulation and Shortages	
more people live on boats	conserve natural resources
emigrate to other planets	increase in technology
more skyscrapers	more homes built for the elderly
increased birth control	
increased research on sea and space communities	
Physical Changes	
more homeless	food shortages
more jails	shortages of water and other resources
increase in sales of false teeth	more air pollution
earth collapses due to environmental stress	
decrease in available land (fights over land; real estate prices increase)	
Social Changes	
life would be more boring	no wills
those now suffering, suffer forever	no life insurance
more/less crime; murder no longer a crime	larger workforce
nature of war changes	people would take more risks
no doctors	social confusion
increase in procrastination	no more concern about death/disease
no promotions	no more cemeteries
wisdom increases; more higher education	no more capital punishment
time not an important factor	more unemployment
decrease in morals/ethics	
less sorrow	
greater burden on the young	
social security system falls apart	
more mental illness	
no religion	

Question 4: Gravity cut in half	
Remote	**Obvious**
Weight Related Effects:	
trucks/planes can carry heavy loads	physical activity easier
trampolines more exciting	changes in weight
people would eat more/be fatter	things/people would weigh half as much
all athletic records would be broken	things float more easily
sky divers in air longer	steps easier to climb
easier to launch space vehicles	greater hang time/easier to dunk
balloon rides different	increase the height of basketball hoop
have to lift heavier weights to get a good workout	basketball players are better
	can throw things farther
Physical Effects	
ceilings need to be higher	orbit/path of the earth would change
people become weaker	more tidal waves
baseball fields made larger in size	tides would change
shoes would last longer	people would grow taller
flying things could have weaker wings	can't go down a hill as fast
people would have weaker bones	atmosphere would change
Miscellaneous	
fishing would be different	Newton's law wrong
moon farther from earth	physics tests need revision
	hair would be higher

Item 5 - No use of arm or hands	
Remote	**Obvious**
Solutions	
new kinds of shoes required	legs/feet become stronger
new machines made	people develop their feet to become versatile
robotics made more useful	mouth used for lifting
voice activation for machinery	write with feet
lower push buttons	car controls changed
door knobs become door pedals	use feet for sign language
Social Changes	
war/battle changed drastically	no handshakes or salutes or clapping
fewer thefts	soccer would become most popular sport
social practices change	can't hug people
everyone walks like a plebe	less fighting
no wearing shoes	college football obsolete
marriage rate down	no surgeons
styles change, beards, no make-up all in style	
Other Changes	
dentists get more work	manicure salons go out of business
people have smaller chests larger legs	no rings
can't test fruit at grocery	can't write with hands
	no zippers
	no guns
	no boxing
	some sports no longer possible, bowling, baseball, basketball, volleyball...

Consequences Scale

Scoring

Rating scales for the Consequences test reflect the original dimensions from the Guilford and Guilford scoring system.

Ideational Fluency - total number of obvious-acceptable responses

- ✔ responses are a direct, immediate result of presented situation
- ✔ responses are commonly associated functions of the presented situation
- ✔ responses are typical or obvious results of the presented situation

---

Originality - total number of remote-acceptable responses

- ✔ novel and imaginative consequences
- ✔ state more than is obvious in the problem
- ✔ results are not an obvious, direct result of the presented situation
- ✔ results are distant temporally or geographically from presented situation
- ✔ results are a specific substitute, or a new system
- ✔ results are a specific way of adjusting to the change

Note: Due to the open-ended, subjective nature of this measure, there is no way to identify all potential responses; thus, scoring will remain, to some degree, subjective.

Note: If a responses can be interpreted in two different ways, such that the rating would change (i.e., could be rated as remote or obvious depending on interpretation); or you group two (or more) responses as duplicates, one original and one remote response, classify the response as remote. (Give the respondent the benefit of the more "creative" rating.)

Consequences Scale (continued)

Unacceptable Responses

Unacceptable Responses are defined as irrelevant and/or duplicate responses, and are not included in the scoring of the test.

Irrelevant Responses

- ✔ Responses not germane to the presented situation
- ✔ Response is so brief that it is uninterpretable and unscoreable
- ✔ Sarcastic responses

> *Note:*
> - ☛ Irrelevant responses may appear germane upon a cursory look, but are not related to the situation when considered more closely.
>
> ***Having said that***:
> - ☛ Be careful not to prematurely eliminate acceptable remote responses (thinking they are not germane to the question.)

Duplicate Responses

- ✔ Repeats one of the example responses
- ✔ Rewording of a previous idea
- ✔ Set of responses from a single category of responses

> *Note:*
> - ☛ Duplicate responses are related by a common theme or idea. To be considered, the responses should be distinct ideas or thoughts.
> - ☛ Error on the conservative side when combining items as duplicate.
> - ☛ Always double check for duplication of the examples provided.

Alternate Headlines

Alternate Headlines Scoring Process

Each of the ten headlines will be rated on three scales: meaning, presentation, and creativity. Two scores will be generated for analysis. Writing Skill will be the mean of the raters' mean presentation and meaning ratings. Creativity will be the mean of the raters' mean creativity ratings.

Meaning: The extent to which the rewritten headline preserves the inherent meaning of the original headline.

0 *Blank*
1 *Rewrite does not reflect meaning of the original*
2 *Rewrite partially reflects meaning of original*
3 *Rewrite reflects meaning of original*
CR *Cannot rate because response cannot be taken seriously*

If an item is rated "0" or "CR"on Meaning, then do not rate on other two scales. If five or more of the ten items are rated "0" or "CR," then do not score the instrument.

Presentation: The extent to which the rewritten headline is grammatically correct (e.g., agreement of subject and verb) and is written in headline format. Effectively formatted headlines avoid past tense and do not reveal details or extraneous information about the article (e.g., specific names).

1 *Headline is not grammatically correct AND (uses past tense OR provides extraneous information.*
2 *Headline is not grammatically correct OR uses past tense AND provides extraneous information.*
3 *Headline is grammatically correct but is written in past tense AND provides extraneous information.*
4 *Headline is grammatically correct but is written in past tense OR provides extraneous information.*
5 *Headline is grammatically correct, avoids past tense, AND contains only essential information.*

Creativity: The extent to which the rewritten headline is imaginative (e.g., catchy, amusing, emotional), uses words that are different from the original headline, and reflects extrapolation from the original meaning.

1 *Rewrite shows no imagination and relies mostly on words from the original*
2
3 *Rewrite shows some novelty and does not over-rely on words from the original*
4
5 *Rewrite is highly imaginative and most words are different from original*

Alternate Headlines (continued)

Changes to scales in Task 1 MFR

- ✔ Meaning scale changed to specify "fundamental" meaning
- ✔ Change presentation scale because the original headlines are not of equivalent quality

---

Meaning Scale

- ✔ Don't penalize for adding information
- ✔ Don't penalize for excluding trivial details (e.g., specifying "rural" highway)

The following are required (or optional) for a high meaning score:
Item 1: Specify a drowning; and that "other" person does **not** survive. Only need to imply a significant other or loved one.
Item 2: Crash over ocean and number of persons are optional
Item 3: Specify that grandmother survives more than one crash and grandma is in plane crash. Desert is optional.
Item 4: Specify that stealing is during test drive. "Duping" car dealer is optional.
Item 5: Note that alone is different than lonesome.
Item 6: Optional to note state and rural descriptors.
Item 7: Specify a police chase, optional to mention what crashes.
Item 8: Specify correlation between low income and inadequate attention, but optional to mention the study.
Item 9: Specify the learning is rewarding concept, and that programs are during the summer.
Item 10: Optional to mention research team.

Presentation Scale

- ✔ Don't score for "catchiness" as this is covered in the creativity scale
- ✔ The average score should be about a 3.

Creativity Scale

- ✔ Might think of this as a "thinking of the audience" scale -- Is respondent showing consideration of the audience by trying to capture their attention or amuse them?

Military Scenarios

Military Scenarios Scoring Process

Each of the two scenarios will be rated on Problem Construction. Ratings will be combined across scenarios and raters to yield a single Problem Construction score.

Problem Construction: The extent to which the examinee effectively frames the overall problem by (a) identifying all the contributing factors that are implied by the scenario description, (b) describing the elements associated with these contributing factors, (c) demonstrating a problem-solving perspective by discussing specifically how these factors and factor elements bear on the problem, and (d) identifying and discussing contributing factors that are not specifically identified in the scenario. Together, these elements assess the breadth (comprehensiveness) and depth of the examinee's response.

CR	*Response is not serious and cannot be rated*
0	*Incomplete or blank responses*
1	*Response focuses on some of the considerations explicitly cited in the scenario description without detailing elements of those considerations*
2	
3	*Response identifies all the key considerations explicitly cited in the scenario description, without discussing the relevance those considerations or identifies a relevant consideration not cited in the scenario*
4	
5	*Response identifies all the key considerations explicitly cited in the scenario description and provides an explicit discussion of how these considerations relate to the solving of the problem or identifies considerations not cited in the scenario.*
6	
7	*Response identifies all the key considerations explicitly cited in the scenario description, as well as considerations not cited in the scenario and provides an explicit discussion of how these considerations relate to the solving of the problem*

Scenario One Explicitly Cited Considerations: Enemy movements, Troops (tired, short-handed, inexperienced replacements), Support options

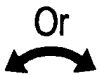
Scenario Two Explicitly Cited Considerations: Influence with Governor, Productivity, Morale, Budget

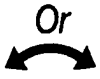
Military Scenarios (continued)

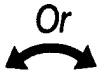
Considerations in scoring Military Scenarios

- ✔ Base rating on entirety of response (i.e., to all that is written in response to the 3 questions posed)
- ✔ Focus on the big picture of the breadth and depth of the response
- ✔ If response doesn't cover all explicitly cited considerations, it should not get a rating of higher than 4.
- ✔ Don't be surprised if ratings to Scenario #1 tend to be lower than those in Scenario #2.
- ✔ If response refers to Army terms with which you are unfamiliar, seek clarification.

Assigned Score	Nature of Response Provided			
	Identification of Cited Consideration	Discussion of Cited Consideration	Identification of Outside Consideration	Discussion of Outside Consideration
1	Some	None	None	None
2	Some Plus	None	None	None
3	All	None	One	None
4	All	Some	One Plus	None
5	All	Pretty Full	Some	None
6	All	Pretty Full	Some	Some
7	All	Pretty Full	Some	Pretty Full







Appendix C - Prescored Sample Responses

Page

Prescored Alternate Headline Responses

	ID# 1061			ID# 1138			ID# 1248			ID# 1345			ID# 1382		
	M	P	C	M	P	C	M	P	C	M	P	C	M	P	C
Headline 1	2	2	3	3	2	3	2	3	2	3	2	1	2	3	3
Headline 2	3	3	3	3	2	3	3	4	2	2	3	1	2	3	1
Headline 3	3	2	3	3	2	3	2	3	3	3	3	1	3	3	3
Headline 4	2	3	4	3	3	2	1	2	3	3	3	3	2	2	3
Headline 5	2	3	4	3	2	2	3	2	4	3	2	3	3	4	3
Headline 6	2	4	5	3	2	2	1	4	4	3	3	2	3	3	3
Headline 7	3	2	2	3	2	2	2	3	4	3	2	2	2	3	2
Headline 8	2	2	3	3	2	2	3	2	2	3	2	2	3	3	3
Headline 9	3	3	4	3	2	2	2	4	5	3	2	1	3	2	2
Headline 10	3	2	2	3	2	1	2	3	3	3	3	1	2	3	3

SSN _____

ALTERNATE HEADLINES

The following test provides you with 10 headlines from major news clippings. Please re each headline to say the same thing using different words. Remember, your rewrites shou in headline format.

Allow yourself one minute for each "rewrite."

Suggested Completion Time: 10 minutes

Sample Headline: "FAMILY PET STALKS GUNMAN."

Sample Rewrite: "FIDO SAVES THE DAY: ROBBERY THWARTED."

1. Headline: "MAN DROWNS IN VAIN EFFORT TO SAVE FIANCEE."

   Rewrite: Hero dies swimming to save Fia

2. Headline: "PLANES COLLIDE OVER OCEAN, KILLING THREE."

   Rewrite: " Three Killed in Air Crash above Pac

3. Headline: "GRANDMOTHER SURVIVES ANOTHER DESERT JET CRASH

   Rewrite: Grandma OK After Plane L
   Desert Dust . . . AGAIN

4. Headline: "CAR THIEVES ARE DUPING SELLERS BY 'TESTING' AND NOT RETURNING."

   Rewrite: " Test NOW- Don't PAY LATER"

1061

5. Headline: "MORE COMMUTERS ARE TRAVELING ALONE."

Rewrite: Lonesome commuters on the Rise

6. Headline: "STATE INMATES KEEP RURAL HIGHWAYS CLEAN."

Rewrite: Prisoners Police Public Places

7. Headline: "AUTO BEING CHASED BY POLICE HITS VAN."

Rewrite: Police chase Car into VAN

8. Headline: "STUDY FINDS LOW-INCOME YOUNGSTERS GET INADEQUATE ATTENTION."

Rewrite: Poor Youth need more help, Study Says.

9. Headline: "SUMMER PROGRAMS MAKE LEARNING REWARDING."

Rewrite: ~~It PAYS to~~ Summer Studies PAY off.

10. Headline: "RESEARCH TEAM FINDS VACCINE THAT COULD SAVE MILLIONS."

Rewrite: Millions will live - New Vaccine Found.

1061 9

SSN _____

ALTERNATE HEADLINES

The following test provides you with 10 headlines from major news clippings. Please rewrite each headline to say the same thing using different words. Remember, your rewrites should be in headline format.

Allow yourself one minute for each "rewrite."

Suggested Completion Time: 10 minutes

Sample Headline: "FAMILY PET STALKS GUNMAN."

Sample Rewrite: "FIDO SAVES THE DAY: ROBBERY THWARTED."

1. Headline: "MAN DROWNS IN VAIN EFFORT TO SAVE FIANCEE."

   Rewrite: ROBERT BROWN DROWNED TODAY WHILE ATTEMPTING TO SAVE HIS FIANCE SHANA HARPER.

2. Headline: "PLANES COLLIDE OVER OCEAN, KILLING THREE."

   Rewrite: THREE USAF PILOTS WERE KILLED TODAY WHEN THEIR PLANES COLLIDED DURING A TRAINING MISSION IN THE ATLANTIC OCEAN NEAR NORFOLK.

3. Headline: "GRANDMOTHER SURVIVES ANOTHER DESERT JET CRASH."

   Rewrite: A GRANDMOTHER OF THREE SURVIVED HER FOURTH DESERT JET CRASH AND VOWED NEVER TO FLY AGAIN.

4. Headline: "CAR THIEVES ARE DUPING SELLERS BY 'TESTING' AND NOT RETURNING."

   Rewrite: CAR THIEVES ARE STEALING CARS DURING TEST DRIVES.

5. Headline: "MORE COMMUTERS ARE TRAVELING ALONE."

Rewrite: AGAINST THE NATIONAL CONSERVATION DESIRE, MORE COMMUTERS ARE TRAVELLING ALONE.

6. Headline: "STATE INMATES KEEP RURAL HIGHWAYS CLEAN."

Rewrite: STATE INMATES DO THEIR PART TO HELP KEEP OUR RURAL HIGHWAYS CLEAN.

7. Headline: "AUTO BEING CHASED BY POLICE HITS VAN."

Rewrite: VAN DEMOLISHED BY AUTO BEING CHASED BY POLICE AT HIGH SPEEDS.

8. Headline: "STUDY FINDS LOW-INCOME YOUNGSTERS GET INADEQUATE ATTENTION."

Rewrite: A NATIONAL STUDY FOUND THAT THE LOWER THE FAMILIES INCOME, THE LESS ATTENTION IS GIVEN TO YOUNGSTERS

9. Headline: "SUMMER PROGRAMS MAKE LEARNING REWARDING."

Rewrite: THE GOAL OF A REWARDING LEARNING EXPERIENCE IS ATTAINED BY SUMMER PROGRAMS.

10. Headline: "RESEARCH TEAM FINDS VACCINE THAT COULD SAVE MILLIONS."

Rewrite: MILLIONS OF PEOPLE COULD BE SAVED BY A RESEARCH TEAM THAT JUST DEVELOPED A VACCINE.

SSN _____

ALTERNATE HEADLINES

The following test provides you with 10 headlines from major news clippings. Please rewrite each headline to say the same thing using different words. Remember, your rewrites should be in headline format.

Allow yourself one minute for each "rewrite."

Suggested Completion Time: 10 minutes

Sample Headline: "FAMILY PET STALKS GUNMAN."

Sample Rewrite: "FIDO SAVES THE DAY: ROBBERY THWARTED."

1. Headline: "MAN DROWNS IN VAIN EFFORT TO SAVE FIANCEE."

   Rewrite: MAN DROWNS TO SAVE LOVED ONE.

2. Headline: "PLANES COLLIDE OVER OCEAN, KILLING THREE."

   Rewrite: AERIAL COLLISION KILLS THREE

3. Headline: "GRANDMOTHER SURVIVES ANOTHER DESERT JET CRASH."

   Rewrite: GRANNY CRASHES AGAIN

4. Headline: "CAR THIEVES ARE DUPING SELLERS BY 'TESTING' AND NOT RETURNING."

   Rewrite: THE ULTIMATE TEST DRIVE

5. Headline: "MORE COMMUTERS ARE TRAVELING ALONE."

Rewrite: TO HELL WITH CARPOOLING.

6. Headline: "STATE INMATES KEEP RURAL HIGHWAYS CLEAN."

Rewrite: CONVICTS CLEAN UP THEIR ACT.

7. Headline: "AUTO BEING CHASED BY POLICE HITS VAN."

Rewrite: HOT PURSUIT ENDS IN TWENTY

8. Headline: "STUDY FINDS LOW-INCOME YOUNGSTERS GET INADEQUATE ATTENTION."

Rewrite: STUDY SHOWS POOR CHILDREN RECEIVE INADEQUATE ATTENTION

9. Headline: "SUMMER PROGRAMS MAKE LEARNING REWARDING."

Rewrite: SUMMER STUDY IS FUN IN THE SUN

10. Headline: "RESEARCH TEAM FINDS VACCINE THAT COULD SAVE MILLIONS."

Rewrite: VACCINE CURES AIDS

SSN: _____

ALTERNATE HEADLINES

The following test provides you with 10 headlines from major news clippings. Please rewrit each headline to say the same thing using different words. Remember, your rewrites should b in headline format.

Allow yourself one minute for each "rewrite."

Suggested Completion Time: 10 minutes

Sample Headline: "FAMILY PET STALKS GUNMAN."

Sample Rewrite: "FIDO SAVES THE DAY: ROBBERY THWARTED."

1. Headline: "MAN DROWNS IN VAIN EFFORT TO SAVE FIANCEE."

   Rewrite: HEROIC EFFORTS TO SAVE FIANCEE WERE IN VAIN : HE DROWNED.

2. Headline: "PLANES COLLIDE OVER OCEAN, KILLING THREE."

   Rewrite: 3 KILLED IN PLANE COLLISION

3. Headline: "GRANDMOTHER SURVIVES ANOTHER DESERT JET CRASH."

   Rewrite: JET CRASH IN DESERT : GRANDMA SURVIVES AT

4. Headline: "CAR THIEVES ARE DUPING SELLERS BY 'TESTING' AND NOT RETURNING."

   Rewrite: CAR SELLERS BEWARE — THIEVES "TEST" CARS AND DON'T RETURN!

5. Headline: "MORE COMMUTERS ARE TRAVELING ALONE."

Rewrite: WHY CAR POOL WHEN MOST COMMUTERS DO IT ALONE.

6. Headline: "STATE INMATES KEEP RURAL HIGHWAYS CLEAN."

Rewrite: COMMUNITY SERVICE BY INMATES — KEEPING YOUR RURAL HIGHWAYS CLEAN.

7. Headline: "AUTO BEING CHASED BY POLICE HITS VAN."

Rewrite: POLICE CHASE AUTO - AUTO COLLIDES INTO VAN!

8. Headline: "STUDY FINDS LOW-INCOME YOUNGSTERS GET INADEQUATE ATTENTION."

Rewrite: LOW INCOME KIDS RECEIVE INADEQUATE ATTENTION ACCORDING TO A RECENT STUDY.

9. Headline: "SUMMER PROGRAMS MAKE LEARNING REWARDING."

Rewrite: LEARNING CAN BE REWARDING THROUGH SUMMER PROGRAMS.

10. Headline: "RESEARCH TEAM FINDS VACCINE THAT COULD SAVE MILLIONS."

Rewrite: VACCINE FOUND THAT COULD SAVE MILLIONS OF PEOPLE!

SSN: _____

ALTERNATE HEADLINES

The following test provides you with 10 headlines from major news clippings. Please rewrite each headline to say the same thing using different words. Remember, your rewrites should be in headline format.

Allow yourself one minute for each "rewrite."

Suggested Completion Time: 10 minutes

Sample Headline: "FAMILY PET STALKS GUNMAN."

Sample Rewrite: "FIDO SAVES THE DAY: ROBBERY THWARTED."

1. Headline: "MAN DROWNS IN VAIN EFFORT TO SAVE FIANCEE."

   Rewrite: Hubby-to-Be Drowns for Fiancee.

2. Headline: "PLANES COLLIDE OVER OCEAN, KILLING THREE."

   Rewrite: Three Killed in Plane Collision

3. Headline: "GRANDMOTHER SURVIVES ANOTHER DESERT JET CRASH."

   Rewrite: "She Does It Again Grandma survives jet crash"

4. Headline: "CAR THIEVES ARE DUPING SELLERS BY 'TESTING' AND NOT RETURNING."

   Rewrite: "Free Test Drive for Free Cars"

5. Headline: "MORE COMMUTERS ARE TRAVELING ALONE."

Rewrite: Commuters Go Solo.

6. Headline: "STATE INMATES KEEP RURAL HIGHWAYS CLEAN."

Rewrite: Cons Clean Highways

7. Headline: "AUTO BEING CHASED BY POLICE HITS VAN."

Rewrite: CHASED CAR HITS VAN

8. Headline: "STUDY FINDS LOW-INCOME YOUNGSTERS GET INADEQUATE ATTENTION."

Rewrite: Low Income Tots: Low Income Athletes

9. Headline: "SUMMER PROGRAMS MAKE LEARNING REWARDING."

Rewrite: Summer School for Better Learning

10. Headline: "RESEARCH TEAM FINDS VACCINE THAT COULD SAVE MILLIONS."

Rewrite: Wonder Vaccine Found

Prescored Consequences Responses

	World Covered With Water		Loss of Reading and Writing		No Death		Gravity Cut in Half		No Use of Arms or Hands	
ID Number	Obvious	Remote	Obvious	Remote	Obvious	Remote	Obvious	Remote	Obvious	Remote
1061	2	3	2	1	0	2	0	2	1	2
1138	3	1	0	3	0	0	1	3	1	0
1248	4	2	4	3	2	0	1	0	0	0
1345	3	2	4	5	4	3	3	3	0	0
1382	3	0	2	6	1	4	1	4	2	0

Breakdown of Consequences Scores for Respondent 1061

World Covered With Water

1. Remote
2. Remote
3. Obvious
4. Duplicate Response of #1
5. Obvious
6. Remote

Loss of Reading and Writing

1. Obvious
2. Obvious
3. Remote

No Death

1. Remote
2. Duplicate Response of #1
3. Duplicate Response of #1
4. Remote
5. Duplicate Response of #4

Gravity Cut in Half

1. Remote
2. Remote

No Use of Arms or Hands

1. Remote
2. Remote
3. Obvious

Breakdown of Consequences Scores for Respondent 1138

World Covered With Water

1. Obvious
2. Duplicate Response of #1
3. Obvious
4. Obvious
5. Remote

Loss of Reading and Writing

1. Cancelled out due to remote duplicate response in #2
2. Remote (Cancels out the obvious duplicate response in #1)
3. Irrelevant Response
4. Remote
5. Remote

No Death

1. Remote
2. Not relevant with the question at hand; irrelevant response
3. Not relevant with the question at hand; irrelevant response
4. Not relevant with the question at hand; irrelevant response
5. Remote

Gravity Cut in Half

1. Obvious
2. Remote
3. Duplicate Response of #2
4. Remote
5. Remote

No Use of Arms or Hands

1. Irrelevant Response
2. Obvious
3. Duplicate Response of #2
4. Irrelevant Response

Breakdown of Consequences Scores for Respondent 1248

World Covered With Water

1. Irrelevant Response--provided as a sample response
2. Remote
3. Obvious
4. Duplicate Response of #1 (an irrelevant response)
5. Obvious
6. Duplicate Response of #5
7. Duplicate Response of #5
8. Remote
9. Obvious
10. Obvious

Loss of Reading and Writing

1. Obvious
2. Remote
3. Obvious
4. Obvious
5. Remote
6. Obvious
7. Remote

No Death

1. Irrelevant Response--provided as sample response
2. Obvious
3. Obvious

Gravity Cut in Half

1. Obvious
2. Irrelevant Response--Incomplete

No Use of Arms or Hands

Breakdown of Consequences Scores for Respondent 1345

World Covered With Water

1. Obvious
2. Remote
3. Duplicate Response of #1
4. Remote
5. Irrelevant Response--duplicate of sample response
6. Obvious
7. Obvious

Loss of Reading and Writing

1. Remote
2. Remote
3. Obvious
4. Duplicate Response of #3
5. Obvious
6. Remote
7. Obvious
8. Obvious
9. Remote
10. Remote
11. Irrelevant Response

No Death

1. Obvious
2. Obvious
3. Irrelevant Response
4. Remote
5. Obvious
6. Remote
7. Remote
8. Irrelevant Response
9. Obvious
10. Duplicate Response of #9

Breakdown of Consequences Scores for Respondent 1345 (continued)

Gravity Cut in Half

1. Obvious
2. Obvious
3. Remote
4. Obvious
5. Duplicate Response of #4
6. Remote
7. Remote

No Use of Arms or Hands

1. Irrelevant Response

Breakdown of Consequences Scores for Respondent 1382

World Covered With Water

1. Irrelevant Response--Duplicate of Sample Response
2. Obvious
3. Duplicate Response of #2
4. Irrelevant Response--Duplicate of Sample Response
5. Obvious
6. Obvious

Loss of Reading and Writing

1. Obvious
2. Duplicate Response of #1
3. Remote
4. Remote
5. Remote
6. Obvious
7. Remote
8. Irrelevant Response--Braille is a type of reading
9. Duplicate Response of #7
10. Remote
11. Duplicate Response of #10

No Death

1. Remote
2. Obvious
3. Remote
4. Remote
5. Remote

Gravity Cut in Half

1. Obvious
2. Cancelled out due to remote duplicate response in #3
3. Remote (Cancels out the obvious duplicate response in #2)
4. Remote
5. Remote
6. Remote

Breakdown of Consequences Scores for Respondent 1382 (continued)

No Use of Arms or Hands

1. Obvious
2. Obvious

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water, except for a few of the highest mountain peaks?

Sample Responses:

a. Everyone will move to mountain peaks
b. Increased sale of boats
c. Business failure
d. Panic

Start Time _____

1. People would prepare to meet their maker
2. People would party alot for 3 months
3. Learn to tread water or swim
4. Soul-searching
5. Research on how to stop the flood
6. Travel to Space
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

1061 3

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if everyone suddenly lost the ability to read and write?

Sample Responses

a. No newspapers or magazines
b. No libraries
c. No mail or letters
d. T.V. sales increase

Start Time _____

1. More talking
2. More watching TV
3. languages modified
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

1061 4

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if human life continued on earth without death?

Sample Responses

a. Overpopulation
b. More old people
c. Housing shortage
d. No more funerals

Start Time _____

1. Space travel _____
2. Space Stations _____
3. Moon colonies _____
4. Live on water _____
5. Live under water _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

1061 5

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if the force of gravity was suddenly cut in half?

Sample Responses
a. Jump higher
b. More accidents
c. Less effort to work
d. Easier to lift things

Start Time _____

1. Space travel easier
2. modification to construction standards
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

1061 6

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if suddenly no one could use their arms or hands?

Sample Responses

a. Learn to use feet more
b. No need for gloves
c. Clothing would be changed
d. Couldn't drive cars

Start Time _____

1. fewer accidents with firearms
2. healthy bodies - more walking
3. Transportation would change but
4. we would still be able to fly planes,
5. drive cars, trains, etc.
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

1061 7

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water, except for a few of the highest mountain peaks?

Sample Responses:
a. Everyone will move to mountain peaks
b. Increased sale of boats
c. Business failure
d. Panic

Start Time _____

1. MANY PEOPLE WOULD DIE
2. HUMAN CIVILIZATION WOULD BE IN JEPOARDY
3. ANIMAL POPULATIONS WOULD BE DECIMATED
4. SEVERE PANDEMONIUM IN POPULATED AREAS
5. ANARCHY WOULD TAKE PLACE
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.

Stop Time _____

3

1138

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if everyone suddenly lost the ability to read and write?

Sample Responses

a. No newspapers or magazines
b. No libraries
c. No mail or letters
d. T.V. sales increase

Start Time _____

1. NO BASIS FOR WRITTEN HISTORY
2. HISTORY WOULD BE PASSED DOWN IN STORIES
3. NO MORE PHD's → GOOD!! Rude!
4. LIFE WOULD BE MORE BASIC.
5. LEARNING WOULD BE DONE HANDS ON.
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if human life continued on earth without death?

Sample Responses
a. Overpopulation
b. More old people
c. Housing shortage
d. No more funerals

Start Time _____

1. MORE WARS TO CONQUER "LIVING SPACE"
2. GENOCIDE
3. MORE KILLINGS AT RANDOM.
4. MORE SUICIDES OF OLD PEOPLE WHEN THEY WERE OF LITTLE USE.
5. VERY LITTLE SOCIETAL CHANGE RESISTANCE TO NEW IDEAS
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if the force of gravity was suddenly cut in half?

Sample Responses
a. Jump higher
b. More accidents
c. Less effort to work
d. Easier to lift things

Start Time _____

1. LESS RAIN FOR PLANTS TO SURVIVE.
2. CRIMINALS ABLE TO TAKE MORE RISKS, MORE CRIME.
3. DAREDEVILS TAKING MORE RISKS, MORE OF THEM
4. OVERALL CONDITIONING OF FOLKS DOWN.
5. NEW STANDARDS, STADIUMS, ETC FOR ATHLETIC EVENTS.
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if suddenly no one could use their arms or hands?

Sample Responses
a. Learn to use feet more
b. No need for gloves
c. Clothing would be changed
d. Couldn't drive cars

Start Time _____

1. MANY PEOPLE WOULD DIE.
2. SOME WOULD ADAPT & CONTINUE TO SURVIVE.
3. LAWS OF NATURE WOULD PREVAIL, ONLY THE STRONG SURVIVE.
4. LIFE WOULD BE MORE BASIC, NO LUXURIES.
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water, except for a few of the highest mountain peaks?

Sample Responses:
a. Everyone will move to mountain peaks
b. Increased sale of boats
c. Business failure
d. Panic

Start Time _____

1. PANIC
2. VIOLENCE
3. PEOPLE MOVING TO MOUNTAINS
4. CHAOS
5. BUILDING RAFTS
6. BUYING BOATS
7. BUILDING BOATS
8. SUICIDE
9. FIGHTING HOMES
10. USE OF SUBMARINES
11.
12.
13.
14.
15.
16.

Stop Time _____

3

1248

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if everyone suddenly lost the ability to read and write?

Sample Responses
a. No newspapers or magazines
b. No libraries
c. No mail or letters
d. T.V. sales increase

Start Time _____

1. LACK OF WRITTEN COMMUNICATION
2. MORE TREES SAVED
3. LESS PAPER WASTE
4. I WOULD NOT BE TAKING THIS SURVEY
5. CONFLICT BETWEEN PEOPLE WHO SPEAK DIFF. LANGUAGES
6. PEOPLE WOULD HAVE TO VERBALLY TEACH CHILDREN THEIR LANGUAGE
7. DECREASED VOCABULARY
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if human life continued on earth without death?

Sample Responses
a. Overpopulation
b. More old people
c. Housing shortage
d. No more funerals

Start Time _____

1. OVERPOPULATION
2. DEPLETION OF RESOURCES
3. WORK BECOMES INSIGNIFICANT
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

5

1248

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if the force of gravity was suddenly cut in half?

Sample Responses
a. Jump higher
b. More accidents
c. Less effort to work
d. Easier to lift things

Start Time _____

1. IT WOULD BE EASIER TO DO THINGS
2. AIRCRAFT WOULD
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if suddenly no one could use their arms or hands?

Sample Responses

a. Learn to use feet more
b. No need for gloves
c. Clothing would be changed
d. Couldn't drive cars

Start Time _____

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water, except for a few of the highest mountain peaks?

Sample Responses:
a. Everyone will move to mountain peaks
b. Increased sale of boats
c. Business failure
d. Panic

Start Time _____

1. BUILD SUBMARINES
2. PRAY
3. BUILD MORE SHIPS
4. MASS SUICIDE
5. MASS HYSTERIA
6. WORK FOR MOUNTAIN PEAKS
7. REAL ESTATE PRICES WOULD DECLINE
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if everyone suddenly lost the ability to read and write?

Sample Responses
a. No newspapers or magazines
b. No libraries
c. No mail or letters
d. T.V. sales increase

Start Time _____

1. COMMUNICATION LOSS LEADS TO DEGENERATE ACTIVITY
2. BUSINESSES IN PUBLISHING WOULD SUFFER
3. VIDEO SALES WOULD INCREASE
4. VIDEO CAMCORDERS WOULD INCREASE IN DEMAND
5. TELEPHONE WOULD BE USED MORE THAN SURE
6. MILITARY OPERATIONS WOULD SUFFER
7. WOULDN'T BE DOING THIS SURVEY THIS WAY
8. VERBAL SKILLS WOULD BE A COMMODITY
9. COLLEGES WOULD HAVE TO REPROGRAM ✓
10. COMPUTER WORLD WOULD SUFFER
11. IT'S HARD TO SAY
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if human life continued on earth without death?

Sample Responses

a. Overpopulation
b. More old people
c. Housing shortage
d. No more funerals

Start Time _____

1. GLOBAL UNREST
2. HOW TO FEED THE WORLD
3. MEDICINES
4. SOCIAL SECURITY WOULD BROKE
5. UNEMPLOYMENT
6. MILLENNIUM IS AT HAND, NO WORRIES
7. A NEW WORLD ORDER — NO WORDS OR RUMORS
8. INCREASE IN PROPERTY —
9. HOUSING PROBLEMS — NO PLACE TO PUT THEM
10. SPACE PROBLEMS
11.
12.
13.
14.
15.
16.

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if the force of gravity was suddenly cut in half?

Sample Responses
a. Jump higher
b. More accidents
c. Less effort to work
d. Easier to lift things

Start Time _____

1. WOULD BE SUPERPEOPLE
2. I HAVE TO TALK TO A PHYSICS PROFESSOR
3. LESS TRUCK DRIVING
4. LESS FATIGUE
5. SOLDIERS COULD CARRY MORE EQUIPMENT
6. ITEMS WOULD HAVE TO BE MADE STRONGER
7. EARTHQUAKES WE WOULD HAVE EROSION
8.
9.
10.
11.
12.
13.
14.
15.
16.

Stop Time _____

6

1345

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if suddenly no one could use their arms or hands?

Sample Responses

a. Learn to use feet more
b. No need for gloves
c. Clothing would be changed
d. Couldn't drive cars

Start Time _____

1. Pray _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

11. _____

12. _____

13. _____

14. _____

15. _____

16. _____

Stop Time _____

7

1345

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if it appeared certain that within three months the entire surface of the earth would be covered with water, except for a few of the highest mountain peaks?

Sample Responses:
a. Everyone will move to mountain peaks
b. Increased sale of boats
c. Business failure
d. Panic

Start Time _____

1. Panic
2. Cheap beach front property
3. Expensive mtn property
4. Increased boat sales
5. Army out of business
6. Swimming lessons big
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

3

1382

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if everyone suddenly lost the ability to read and write?

Sample Responses
a. No newspapers or magazines
b. No libraries
c. No mail or letters
d. T.V. sales increase

Start Time _____

1. More TV watched
2. More movies watched
3. Voice computers etc.
4. More time on recreational activities
5. New education techniques
6. More phone use.
7. Libraries, book stores closed
8. Brail becomes a hit.
9. Newspapers close
10. Less paper production
11. Forestry industry collapses.
12.
13.
14.
15.
16.

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if human life continued on earth without death?

Sample Responses

a. Overpopulation
b. More old people
c. Housing shortage
d. No more funerals

Start Time _____

1. Birth control becomes critical.
2. Not enough food → famine
3. Older people become very bored
4. Crime wave by bored adults.
5. Unpopulated areas developed
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if the force of gravity was suddenly cut in half?

Sample Responses

a. Jump higher
b. More accidents
c. Less effort to work
d. Easier to lift things

Start Time _____

1. Don't trip.
2. Machinery becomes more efficient.
3. Reduced need for oil, gas etc
4. Basketball rim height to 20'
5. Takes longer to get in good physical condition.
6. Most people die - not enough oxygen to cells.
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

6

1382

LIST AS MANY DIFFERENT CONSEQUENCES AS YOU CAN.

What would be the results if suddenly no one could use their arms or hands?

Sample Responses

a. Learn to use feet more
b. No need for gloves
c. Clothing would be changed
d. Couldn't drive cars

Start Time _____

1. Soccer becomes favorit sport.
2. Many traffic accidents.
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____

Stop Time _____

1382 7

Prescored Military Scenarios Responses

ID Number	Scenario 1	Scenario 2
1061	3	5
1138	4	4
1248	3	3
1345	4	3
1382	2	2

SSN: _____

MILITARY SCENARIOS

We are interested in how people approach and solve complicated problems they may encounter. Please read each of the 2 following scenarios carefully. Then answer the questions that follow in the space provided.

Suggested Completion Time: 10 Minutes

SCENARIO #1

You are a corps commander whose forces are tired and short-handed as a result of several weeks of rugged fighting. A significant number of your troops are inexperienced replacements who have recently been assigned to their units. Although you have been given some intelligence indicating that the enemy may be massing troops on your front, you do not consider it a serious threat due to the overall state of the war. You have assured your superiors that you have adequate strength to defend your sector, and have resisted suggestions that your forces be folded into a neighboring command with fresher troops. However, it soon becomes apparent that the enemy is mounting a major effort.

1) If you were placed in this situation, what would be the most important problem for you to address?

Protection of my Force enemy attack

2) What key pieces of information would you need to solve the problem?

Size of enemy force

enemy Location

Probable course of Action (enemy)

3) Are there other problems you would have to consider?

-need fresh troops

- need experienced/ trained troops

-need access to friendly reinforcements

-need rest for some of my units

1061

SCENARIO #2:

Upon retirement from military service you have been appointed head of a large government agency in your home state. Your appointment came about in part from your success and achievements during a distinguished military career; which was well publicized in the local press. You view this as being the first step in what you hope will be a long career of appointed or elected government service. However, for this to happen, you must demonstrate that you can be successful in a politically charged civilian environment.

In recent years, your agency has not been as effective as it might have been in dealing with problems which have now become quite serious and require decisive action. Observers of governmental operations within the state have felt that your predecessor had little influence with the Governor in developing policy in your agency's area of responsibility. As a result of this, the productivity and morale of the employees within your department is very low. In addition, budgetary constraints resulting form lower state revenues will effect your efforts to improve your agency's contribution to the state's welfare.

1) If you were placed in this situation, what would be the most important problem for you to address?

PEOPLE - Morale & Productivity must come up and my people can make that happen, despite budgetary constraints and bad rep.

2) What key pieces of information would you need to solve the problem?

- why little influence in the past?
- How much $ do I need?
- why low productivity & morale

3) Are there other problems you would have to consider?

- Get to know Governor and be responsive to states needs
-

1061 Z8

SSN: [redacted]

MILITARY SCENARIOS

We are interested in how people approach and solve complicated problems they may encounter. Please read each of the 2 following scenarios carefully. Then answer the questions that follow in the space provided.

Suggested Completion Time: 10 Minutes

SCENARIO #1:

You are a corps commander whose forces are tired and short-handed as a result of several weeks of rugged fighting. A significant number of your troops are inexperienced replacements who have recently been assigned to their units. Although you have been given some intelligence indicating that the enemy may be massing troops on your front, you do not consider it a serious threat due to the overall state of the war. You have assured your superiors that you have adequate strength to defend your sector, and have resisted suggestions that your forces be folded into a neighboring command with fresher troops. However, it soon becomes apparent that the enemy is mounting a major effort.

1) If you were placed in this situation, what would be the most important problem for you to address?

I WOULD WANT TO PRESERVE AS MUCH FIGHTING FORCE AS POSSIBLE TO FIGHT ANOTHER DAY

2) What key pieces of information would you need to solve the problem?

INTELLIGENCE ON ENEMY

STATUS OF CLOSE AIR SUPPORT

REINFORCING UNITS STATUS & AVAILABILITY.

3) Are there other problems you would have to consider?

HOW MY SITUATION FITS IN WITH THE OVERALL

CDR'S INTENTIONS OF HOW THE BATTLE WOULD BE FOUGHT.

SCENARIO #2

Upon retirement from military service you have been appointed head of a large government agency in your home state. Your appointment came about in part from your success and achievements during a distinguished military career which was well publicized in the local press. You view this as being the first step in what you hope will be a long career of appointed or elected government service. However, for this to happen, you must demonstrate that you can be successful in a politically charged civilian environment.

In recent years, your agency has not been as effective as it might have been in dealing with problems which have now become quite serious and require decisive action. Observers of governmental operations within the state have felt that your predecessor had little influence with the Governor in developing policy in your agency's area of responsibility. As a result of this, the productivity and morale of the employees within your department is very low. In addition, budgetary constraints resulting form lower state revenues will effect your efforts to improve your agency's contribution to the state's welfare.

1) If you were placed in this simation, what would be the most important problem for you to address?

DEFINE A CLEAR AND EASILY UNDERSTOOD GOAL FOR THE ORGANIZATION TO ATTAIN AND A BASIC ROAD MAP OF HOW WE WERE GOING TO GET THERE

2) What key pieces of information would you need to solve the problem?

- GOVERNOR'S INTENTIONS AND SUPPORT
- COORDINATION WITH OTHER GOV'T AGENCIES AS TO WHAT AREAS THEY COULD ASSIST IN.

3) Are there other problems you would have to consider?

- CURRENT STATUS OF LEGISLATIVE SUPPORT
- ABILITY TO INFLUENCE " "
- GOVERNMENT'S FUTURE BUDGET

SSN: _____

MILITARY SCENARIOS

We are interested in how people approach and solve complicated problems they may encounter. Please read each of the 2 following scenarios carefully. Then answer the questions that follow in the space provided.

Suggested Completion Time: 10 Minutes

SCENARIO #1

You are a corps commander whose forces are tired and short-handed as a result of several weeks of rugged fighting. A significant number of your troops are inexperienced replacements who have recently been assigned to their units. Although you have been given some intelligence indicating that the enemy may be massing troops on your front, you do not consider it a serious threat due to the overall state of the war. You have assured your superiors that you have adequate strength to defend your sector, and have resisted suggestions that your forces be folded into a neighboring command with fresher troops. However, it soon becomes apparent that the enemy is mounting a major effort.

1) If you were placed in this situation, what would be the most important problem for you to address?

PREPARING MY TROOPS FOR AN ATTACK

2) What key pieces of information would you need to solve the problem?

TIME AVAILABLE, SPECIFIC WEAKNESSES IN MY DEFENSES, POSSIBILITY OF ATTACKING THE ENEMY BEFORE HE ATTACKS ME.

3) Are there other problems you would have to consider?

YES, I WOULD HAVE TO CHANGE WHAT I TOLD MY SUPERIORS.

[redacted] 1248

SCENARIO #2

Upon retirement from military service you have been appointed head of a large government agency in your home state. Your appointment came about in part from your success and achievements during a distinguished military career, which was well publicized in the local press. You view this as being the first step in what you hope will be a long career of appointed or elected government service. However, for this to happen, you must demonstrate that you can be successful in a politically charged civilian environment.

In recent years, your agency has not been as effective as it might have been in dealing with problems which have now become quite serious and require decisive action. Observers of governmental operations within the state have felt that your predecessor had little influence with the Governor in developing policy in your agency's area of responsibility. As a result of this, the productivity and morale of the employees within your department is very low. In addition, budgetary constraints resulting form lower state revenues will effect your efforts to improve your agency's contribution to the state's welfare.

1) If you were placed in this situation, what would be the most important problem for you to address?

INFLUENCE WITH THE GOVERNOR IN DEVELOPING POLICY

2) What key pieces of information would you need to solve the problem?

WAYS TO IMPROVE MORALE AND PRODUCTIVITY, INCREASE REVENUES

3) Are there other problems you would have to consider?

POLITICS

SSN: _____

MILITARY SCENARIOS

We are interested in how people approach and solve complicated problems they may encounter. Please read each of the 2 following scenarios carefully. Then answer the questions that follow in the space provided.

Suggested Completion Time: 10 Minutes

SCENARIO #1

You are a corps commander whose forces are tired and short-handed as a result of several weeks of rugged fighting. A significant number of your troops are inexperienced replacements who have recently been assigned to their units. Although you have been given some intelligence indicating that the enemy may be massing troops on your front, you do not consider it a serious threat due to the overall state of the war. You have assured your superiors that you have adequate strength to defend your sector, and have resisted suggestions that your forces be folded into a neighboring command with fresher troops. However, it soon becomes apparent that the enemy is mounting a major effort.

1) If you were placed in this situation, what would be the most important problem for you to address?

Intelligence Preparation of the Battlefield
- DO I "FIGHT WITH OTHER FORCES"
- Where would the enemy attack?

HAVE MY STAFF DO A DECISION MATRIX & COME UP WITH AT LEAST 3 COURSES OF ACTION FOR MY DECISIONS ON THIS PROBLEM.

2) What key pieces of information would you need to solve the problem?

S-2 INTELLIGENCE UPDATE

3) Are there other problems you would have to consider?

SOLDIERS EXPERIENCE LEVEL THAT AREN'T INEXPERIENCED

SCENARIO #2

Upon retirement from military service you have been appointed head of a large government agency in your home state. Your appointment came about in part from your success and achievements during a distinguished military career, which was well publicized in the local press. You view this as being the first step in what you hope will be a long career of appointed or elected government service. However, for this to happen, you must demonstrate that you can be successful in a politically charged civilian environment.

In recent years, your agency has not been as effective as it might have been in dealing with problems which have now become quite serious and require decisive action. Observers of governmental operations within the state have felt that your predecessor had little influence with the Governor in developing policy in your agency's area of responsibility. As a result of this, the productivity and morale of the employees within your department is very low. In addition, budgetary constraints resulting form lower state revenues will effect your efforts to improve your agency's contribution to the state's welfare.

1) If you were placed in this simation, what would be the most important problem for you to address?

TO DETERMINE THE MOST EFFECTIVE WAY TO GET MY AGENCY BACK ON TRACK

2) What key pieces of information would you need to solve the problem?

WHAT IS CURRENT POLICY

WHAT WERE THE OPERATING PROCEDURES BEFORE THE DECLINE

3) Are there other problems you would have to consider?

EMPLOYEE MORALE

RELATIONSHIP WITH GOVERNOR & OFFICIALS

SSN: _____

MILITARY SCENARIOS

We are interested in how people approach and solve complicated problems they may encounter. Please read each of the 2 following scenarios carefully. Then answer the questions that follow in the space provided.

Suggested Completion Time: 10 Minutes

SCENARIO #1

You are a corps commander whose forces are tired and short-handed as a result of several weeks of rugged fighting. A significant number of your troops are inexperienced replacements who have recently been assigned to their units. Although you have been given some intelligence indicating that the enemy may be massing troops on your front, you do not consider it a serious threat due to the overall state of the war. You have assured your superiors that you have adequate strength to defend your sector, and have resisted suggestions that your forces be folded into a neighboring command with fresher troops. However, it soon becomes apparent that the enemy is mounting a major effort.

1) If you were placed in this simation, what would be the most important problem for you to address?

- defensive plan

2) What key pieces of information would you need to solve the problem?

- exact enemy strength
enemy water supply

3) Are there other problems you would have to consider?

logistics

SCENARIO #2:

Upon retirement from military service you have been appointed head of a large government agency in your home state. Your appointment came about in part from your success and achievements during a distinguished military career, which was well publicized in the local press. You view this as being the first step in what you hope will be a long career of appointed or elected government service. However, for this to happen, you must demonstrate that you can be successful in a politically charged civilian environment.

In recent years, your agency has not been as effective as it might have been in dealing with problems which have now become quite serious and require decisive action. Observers of governmental operations within the state have felt that your predecessor had little influence with the Governor in developing policy in your agency's area of responsibility. As a result of this, the productivity and morale of the employees within your department is very low. In addition, budgetary constraints resulting form lower state revenues will effect your efforts to improve your agency's contribution to the state's welfare.

1) If you were placed in this simation, what would be the most important problem for you to address?

working relationship with Governor - get his ear

2) What key pieces of information would you need to solve the problem?

- old policy, old budget

3) Are there other problems you would have to consider?

Based on the scenario, no

Appendix D - Rater Feedback Forms

Consequences

Feedback Form

What problems did you encounter when scoring responses on this measure?

What parts of the measure made it difficult to use?

Were you uncertain about whether responses were obvious or remote?
Were you uncertain about whether or not responses were relevant?
Were you uncertain about whether groups of responses would be considered duplicate?

What parts of the measure made it easy to use?

Would further training in any specific aspect of the measure be helpful for scoring the responses?

Alternate Headlines

Feedback Form

What problems did you encounter when scoring responses on this measure?

What parts of the measure made it difficult to use?

Was any scale (i.e., meaning, presentation, creativity) more confusing or difficult to use?

Was it difficult to make three independent ratings using these scales?

What parts of the measure made it easy to use?

Would further training in any specific aspect of the measure be helpful for scoring the responses?

Military Scenarios

Feedback Form

What problems did you encounter when scoring responses on this measure?

What parts of the measure made it difficult to use?
Did you find the scale difficult to interpret or to use?
Do you think that one scale was sufficient?

What parts of the measure made it easy to use?

Would further training in any specific aspect of the measure be helpful for scoring the responses?

General Feedback Form

Would further training in any specific aspect of the measure be helpful for scoring the responses?

Which scale did you find the easiest to rate? Why?

Which scale had the most variance in the scores (responses)?

Was it difficult to score all of the responses using the same scoring scales? (The responses you scored were drawn from samples that included officers in grades 01 to 06. Did you have to change your perspective or metric to score any of the responses?)

Appendix E - Descriptive Statistics and Interrater Reliability Analysis

CONSEQUENCES: BASIC SCORES

Consequences Obvious Responses								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	2.17	.72	1.79	.65	2.01	.82	1.78	.96
Rater 2	1.77	.63	1.55	.71	1.83	.63	1.85	.77
Rater 3	1.68	.64	1.38	.51	1.51	.61	1.61	.77
	N=75		N=76		N=75		N=75	

Consequences Remote Responses								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	1.70	.68	1.84	.78	1.77	.72	1.84	.98
Rater 2	2.05	.78	2.02	.74	1.99	.92	1.53	.75
Rater 3	1.66	.59	2.26	.92	2.28	.88	1.93	.95
	N=75		N=76		N=75		N=75	

Consequences Obvious Response Reliabilities	
Dataset 1	
3 Raters	2 Raters
.83	.80
	.67
	.82
Dataset 2	
3 Raters	2 Raters
.87	.79
	.86
	.81
Dataset 3	
3 Raters	2 Raters
.82	.71
	.71
	.85
Dataset 4	
3 Raters	2 Raters
.92	.88
	.87
	.92

Consequences Remote Response Reliabilites	
Dataset 1	
3 Raters	2 Raters
.86	.83
	.77
	.82
Dataset 2	
3 Raters	2 Raters
.92	.87
	.90
	.89
Dataset 3	
3 Raters	2 Raters
.92	.86
	.87
	.91
Dataset 4	
3 Raters	2 Raters
.94	.88
	.93
	.91

ALTERNATE HEADLINES: BASIC SCORES

Scale A: Meaning								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	2.38	.35	2.45	.36	2.23	.46	2.26	.41
Rater 2	2.22	.38	2.21	.38	2.48	.38	2.39	.35
Rater 3	2.43	.30	2.36	.36	2.51	.40	2.35	.41
	N=75		N=75		N=73		N=73	

Scale B: Presentation								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	3.08	.42	2.60	.32	2.50	.31	2.75	.28
Rater 2	3.43	.39	2.90	.26	2.95	.18	2.82	.17
Rater 3	2.82	.23	2.54	.33	2.63	.26	2.59	.33
	N=75		N=75		N=73		N=73	

Scale C: Creativity								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	2.93	.55	2.72	.54	2.72	.73	2.67	.51
Rater 2	2.82	.66	2.77	.50	2.93	.55	2.75	.47
Rater 3	2.72	.46	2.36	.61	2.65	.52	2.20	.50
	N=75		N=75		N=73		N=73	

Alternate Headlines Scale A: Meaning
Dataset 1
3 Raters
.93
Dataset 2
3 Raters
.95
Dataset 3
3 Raters
.95
Dataset 4
3 Raters
.96

Alternate Headlines Scale B: Presentation
Dataset 1
3 Raters
.69
Dataset 2
3 Raters
.63
Dataset 3
3 Raters
.57
Dataset 4
3 Raters
.55

Alternate Headlines Scale C: Creativity
Dataset 1
3 Raters
.89
Dataset 2
3 Raters
.90
Dataset 3
3 Raters
.92
Dataset 4
3 Raters
.91

ALTERNATE HEADLINES: COMPOSITE SCORES

Alternate Headlines Creativity								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	2.93	.55	2.72	.54	2.72	.73	2.67	.51
Rater 2	2.82	.66	2.77	.50	2.93	.55	2.75	.47
Rater 3	2.72	.46	2.36	.61	2.65	.52	2.20	.50
	N=75		N=75		N=73		N=73	

Alternate Headlines Written Skill								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	5.45	.47	5.04	.51	4.73	.58	5.00	.45
Rater 2	4.64	.57	5.12	.38	5.43	.40	5.21	.39
Rater 3	5.25	.36	4.91	.46	5.15	.56	4.94	.61
	N=75		N=75		N=73		N=73	

Alternate Headlines Creativity Reliabilities	
Dataset 1	
3 Raters	2 Raters
.89	.89
	.83
	.81
Dataset 2	
3 Raters	2 Raters
.90	.88
	.80
	.88
Dataset 3	
3 Raters	2 Raters
.92	.89
	.86
	.89
Dataset 4	
3 Raters	2 Raters
.91	.91
	.88
	.81

Alternate Headlines Written Skill Reliabilities	
Dataset 1	
3 Raters	2 Raters
.78	.75
	.70
	.64
Dataset 2	
3 Raters	2 Raters
.81	.71
	.76
	.73
Dataset 3	
3 Raters	2 Raters
.90	.84
	.90
	.84
Dataset 4	
3 Raters	2 Raters
.87	.83
	.80
	.81

MILITARY SCENARIOS: BASIC SCORES

Scenario 1								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	3.22	.97	2.67	1.20	3.05	1.00	2.84	.86
Rater 2	3.25	1.03	2.93	1.09	3.73	.90	3.10	1.24
Rater 3	3.76	.97	3.08	1.03	2.98	1.03	2.68	1.01
	N=59		N=60		N=60		N=59	

Scenario 2								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	3.03	1.08	2.95	1.14	3.06	.94	2.98	1.06
Rater 2	3.17	.95	2.80	1.00	3.61	.98	3.20	1.23
Rater 3	3.76	1.01	2.83	.93	3.08	1.09	2.61	1.00
	N=59		N=59		N=60		N=59	

Military Scenarios Scenario 1
Dataset 1
3 Raters
.84
Dataset 2
3 Raters
.87
Dataset 3
3 Raters
.78
Dataset 4
3 Raters
.87

Military Scenarios Scenario 2
Dataset 1
3 Raters
.82
Dataset 2
3 Raters
.85
Dataset 3
3 Raters
.84
Dataset 4
3 Raters
.85

MILITARY SCENARIOS: COMPOSITE SCORES

Military Scenarios Problem Construction (PC)								
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	3.13	.82	2.78	.99	3.06	.77	2.92	.76
Rater 2	3.21	.76	2.84	.83	3.68	.79	3.15	1.10
Rater 3	3.76	.83	2.96	.81	3.03	.81	2.64	.84
	N=59		N=60		N=60		N=59	

Military Scenarios PC Reliabilities	
Dataset 1	
3 Raters	2 Raters
.85	.72
	.79
	.86
Dataset 2	
3 Raters	2 Raters
.90	.87
	.87
	.81
Dataset 3	
3 Raters	2 Raters
.87	.82
	.78
	.86
Dataset 4	
3 Raters	2 Raters
.90	.78
	.86
	.91

Appendix F - Interrater Agreement Analyses

Rater Agreement

Measures of Agreement - Lawlis and Lu		
Alternate Headlines-Item	3-Raters (χ^2)	2-Raters (χ^2)
1. a) Meaning	110	27
b) Presentation	55	11
c) Creativity	55	21
2. a) Meaning	161	87
b) Presentation	25	1
c) Creativity	20	6
3. a) Meaning	143	55
b) Presentation	15	.06
c) Creativity	119	32
4. a) Meaning	94.64	22
b) Presentation	25.33	11
c) Creativity	.11	1
5. a) Meaning	41	5
b) Presentation	39	16
c) Creativity	73	21
6. a) Meaning	134	38
b) Presentation	32	16
c) Creativity	84	39
7. a) Meaning	41	7
b) Presentation	64	13
c) Creativity	39	26
8. a) Meaning	81	20
b) Presentation	32	23
c) Creativity	107	18
9. a) Meaning	88	18
b) Presentation	95	26

Measures of Agreement - Lawlis and Lu		
Alternate Headlines-Item	3-Raters (χ^2)	2-Raters (χ^2)
c) Creativity	161	29
10. a) Meaning	143	51
b) Presentation	39	13
c) Creativity	11	8
Military Scenarios - Item	3-Rater (χ^2)	2-Rater (χ^2)
Scenario 1	151	41
Scenario 2	59	57

Note: χ^2 test > 6.635 (df = 1) is significant to the .01 level
Shaded blocks represent non-significant levels of rater agreement

Appendix G - Internal Consistency Reliability Analyses

Alternate Headlines: Internal Consistency Reliabilities

Dataset 1

	Alpha if Item Deleted		
Item	Scale A (Meaning)	Scale B (Presentation)	Scale C (Creativity)
Headline 1	.6788	.6443	.8275
Headline 2	.6844	.6448	.8217
Headline 3	.6751	.6478	.8015
Headline 4	.6838	.6932	.8179
Headline 5	.7051	.6566	.8302
Headline 6	.6775	.6735	.8290
Headline 7	.6754	.6587	.8274
Headline 8	.6494	.7063	.8177
Headline 9	.6750	.6689	.8133
Headline 10	.6872	.6567	.8184
Overall alpha	.7021	.6888	.8357
N	75	63	63

Dataset 2

	Alpha if Item Deleted		
Item	Scale A (Meaning)	Scale B (Presentation)	Scale C (Creativity)
Headline 1	.6579	.6556	.8730
Headline 2	.6489	.6427	.8684
Headline 3	.6499	.6521	.8631
Headline 4	.6401	.6076	.8693
Headline 5	.6178	.6367	.8806
Headline 6	.6539	.6649	.8701
Headline 7	.6712	.6595	.8651
Headline 8	.6335	.6683	.8661
Headline 9	.6768	.6502	.8661
Headline 10	.6761	.6261	.8699
Overall alpha	.6766	.6709	.8808
N	75	56	56

Alternate Headlines: Internal Consistency Reliabilities (continued)

Dataset 3

	Alpha if Item Deleted		
Item	Scale A (Meaning)	Scale B (Presentation)	Scale C (Creativity)
Headline 1	.7720	.5745	.8788
Headline 2	.7764	.5669	.8834
Headline 3	.7730	.5791	.8734
Headline 4	.7934	.5630	.8876
Headline 5	.7461	.5673	.8756
Headline 6	.7776	.5383	.8760
Headline 7	.7690	.5940	.8817
Headline 8	.7463	.5627	.8725
Headline 9	.7592	.5874	.8765
Headline 10	.7760	.5869	.8707
Overall alpha	.7879	.5982	.8886
N	73	61	61

Dataset 4

	Alpha if Item Deleted		
Item	Scale A (Meaning)	Scale B (Presentation)	Scale C (Creativity)
Headline 1	.8101	.6131	.8532
Headline 2	.7921	.5764	.8504
Headline 3	.8007	.6019	.8505
Headline 4	.8029	.5868	.8479
Headline 5	.7824	.5519	.8495
Headline 6	.7718	.5672	.8421
Headline 7	.7928	.5812	.8419
Headline 8	.7571	.5889	.8454
Headline 9	.7914	.5684	.8426
Headline 10	.7841	.5517	.8477
Overall alpha	.8063	.6047	.8603
N	74	61	61