

Software Support in the Classroom: Help or HINDRANCE

William H. Gwinn
Information Systems and Operations Management Department,
University of North Carolina – Wilmington,
Wilmington, NC, 28403, U.S.A.

Abstract

Few researchers have addressed the question of how information system requirements should be derived. The rapidly changing needs of increasingly complex organizations are pressuring the analyst to rapidly produce information requirements. This means the analyst needs the capability to rapidly acquire, organize and analyze organizational facts from which information requirements are derived. This research concerns the testing of an adaptive analyst support system to assist the novice analyst (student) with the gathering and managing of organizational facts. The experiment investigates the use of a graphical user interface (GUI) tool to help the student analyst perform organizational fact gathering tasks preliminary to information system requirements determination and specification. The experiment results are discussed and conclusions are drawn from the results of the dual tasks facing a novice analyst when a software tool is provided.

Keywords: Dual tasking, Expert Systems, Analyst Assistant, Software Support

1. INTRODUCTION

One of the hardest concepts for Systems Analysis and Design students to master is the process of gathering the organizational facts necessary for the eventual formulation of Information System requirements. All analysts face the task of gathering and organizing organizational facts prior to beginning the analysis steps leading to the formulation of system requirements (Satzinger 2000). Novice analysts, especially students attempting to implement the steps learned in the classroom setting, struggle with the concepts while attempting to apply theory to organizational scenarios. Systems Analysis and Design students during the Fall of 1998 and the Spring of 1999 were provided with a prototype case-based analyst adaptive support system (CAASS) to assist them in performing the fact gathering and organization tasks of systems analysis (Haag 2000). To determine the effectiveness of the CAASS tool, the students were to perform requirement determination tasks with and without the use of the CAASS (Dennis 2000). The purpose of this paper is to review the results of this experiment.

2. TASKING LITERATURE

A number of researchers have studied the psychological aspects of performing multiple tasks and how performance may be degraded by the human information

processing overload that may result from multi-task processing. "Information resources such as computer systems are typically used in tasks to improve task outcomes . . . such results are not always realized" (Collins 1993, p. 18). Collins (1993) contends that tasks that have inconsistent information processing requirements will not be able to develop automatic processes. Tasks should be specifically chosen to present a high-order consistency of information processing requirements so that automatic processing can develop. Students should be asked to make highly similar decisions about similar situations if automatic processing development is desired (Collins 1993). Tasks that do not become automatic take conscious attention and will cause a decrease in performance for individuals faced with multiple tasks (Thorngate 1976).

Schneider and Fisk (1982) found subjects performing dual task processing experiments could complete multiple tasks without performance degradation if they were able to achieve automatic processing. If the processing had not become automatic, significant decrements in performance occurred despite intensive training on the task performance. Collins (1993) reiterated that degradation in dual task performance may be avoided if information technology use can become automatic.

3. CAASS SOFTWARE

The software tool (CAASS) was a prototype Case-Based artificial intelligent program, designed and built by the author, that employed a familiar Microsoft Windows graphical user interface. All of the students participating in the experiment had received courses in COBOL, C++, and Microsoft Access prior to using the CAASS and some students had also completed courses in Microsoft Visual Basic and/or PowerSoft's PowerBuilder. Based on the experience level of the students in the systems analysis class, a minimum level of familiarization with the software tool was believed to be necessary to minimize the impact of multitasking on student performance. The CAASS at start up provided the student analysts with a checklist of fact gathering activities. As the activities were completed, the results were entered into the CAASS and the facts were categorized to facilitate information system requirement formulation and stored in a template structure (see Figure 1 below).

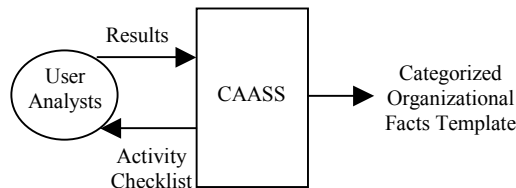


Figure 1. User-CAASS interaction.

4. EXPERIMENT DESIGN

A pilot study using students from the graduate systems analysis class was conducted. The pilot study group pre-tested the fact gathering projects. The projects constituted the pre-test and post-test for the experiment. Experience with this graduate group helped establish the 90-minute period allowed for completion of each test, and clarified the detailed instructions for the use of the CAASS.

Group Selection

In order to provide an adequate number of test subjects on site, undergraduate systems analysis students were used as subjects for the experiment. Seventy-two students in the undergraduate systems analysis course participated in the experimental evaluation of the CAASS system. All students had received instruction in interviewing and fact gathering in their systems analysis course before participating in the CAASS evaluation experiment. Thirty-six subjects were randomly selected from the population of 72 undergraduate systems analysis students to form a control group. The control group performed two requirements determination projects manually. The control group's performance established a baseline level of performance for the pre-test and post-test. A treatment group was formed from the remaining 36 students to use the CAASS and

provide a performance comparison with the baseline control group. The treatment group performed the first project manually and used the CAASS system to perform the second project. See Figure 2 below.

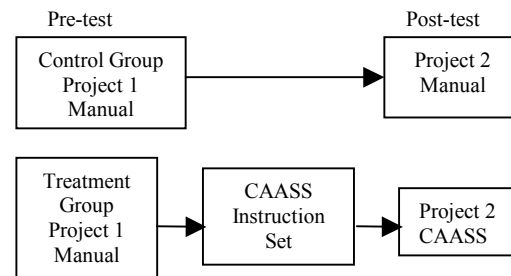


Figure 2. Experiment Layout.

Time demands from class projects, homework, and work schedules restricted the available CAASS testing participation time for each student to three hours. Students were allowed a 90-minute period to complete each test as established in the pilot study. The time constraints on student availability prohibited giving hands on instruction sessions for CAASS usage. As an alternative to hands on instruction sessions, a set of detailed instructions for CAASS operation was provided each CAASS user.

The experimental design layout was a two factor crossed design with repeated measure over the pre-post test factor. This particular design was chosen because "it is the most frequently used design in social sciences research" (Cook 1979, p. 103). It can be used to evaluate equal and non-equal sized groups (Cook 1979; Keppel 1991; Montgomery 1991). A test factor with two levels and a treatment factor of two levels made up the statistical model and layout.

Testing Instrument

The test instrument contained a pre-test level and a post-test level. Each level or case was selected from a workbook by George and Annette Easton (1996). The case workbook is a companion book to the students' systems analysis textbook (Hoffer 1999). The pre-test consisted of a case about Media Technology Services (MTA) at a community college. The post-test was a case about Homeowners of America, a management services firm. The two cases were selected so they would be similar and provide the students with an opportunity to improve their performance by repeating requirements gathering activities on similar cases. Before the experiment was conducted, each case was analyzed by the researcher and had an organizational fact list prepared. The fact lists were validated by comparing the pilot study and experiment student responses to the fact lists. The percentage of reasonable student responses that matched the pre-test fact list was 81.42 percent. The post-test responses exhibited an 88.42 percent match with the fact list. The facts each student identified while performing the pre-post tasks

were compared to the pre-experiment fact lists. The ratio of the number of facts identified by a student on a test to the number of facts in the appropriate pre-experiment fact list produced a performance percentage score for that student. The difference between the control group's pre-test and post-test scores reflected learning from repeating similar tasks. The difference between the tool group's pre-test and post-test scores reflected the learning effect and treatment or tool effect.

Experiment Layout and Model

The two factor crossed experimental design with repeated measure over the test factor produced the layout below (Table 1). The statistical model for this experimental design is:

$$y_{i,j,k} = \mu + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad (1)$$

where $i = 1,2$; $j = 1,2$; and $k = 1,36$. In this model $y_{i,j,k}$ is the observed score for the i th treatment, j th test, and k th subject, μ is the overall mean, τ_i is the effect for the i th level of the treatment factor, β_j is the effect for the j th level of the test factor, $(\tau\beta)_{ij}$ is an interaction term for interaction between test and treatment factors, and ϵ_{ijk} is a random error term (Montgomery 1991). The Analysis of Variance (ANOVA) with interaction was chosen specifically because an interaction was expected.

Table 1. Experimental Design Layout

TREATMENT	SUBJECTS	PRE-TEST (1)	POST-TEST (2)
CONTROL (1)	Subj ₁	score _{1,1,1}	score _{1,2,1}
	Subj ₂	score _{1,1,2}	score _{1,2,2}
	•	•	•
	•	•	•
	Subj ₃₆	score _{1,1,36}	score _{1,2,36}
TOOL (2)	Subj ₃₇	score _{2,1,37}	score _{2,2,37}
	Subj ₃₈	score _{2,1,238}	score _{2,2,38}
	•	•	•
	•	•	•
	Subj ₇₂	score _{2,1,72}	score _{2,2,72}

Both the control group and treatment group performed the pre-test manually. The results from both groups were expected to be very close with no significant difference on the pre-test and a significant difference on the post-test.

Hypothesis Testing

The experiment was designed to allow the posing of hypotheses about the component treatments and tests. The post-test score for the control group was expected to be higher than the pre-test score, reflecting the learning from repeated similar tasks. The post-test score for the

treatment (tool) group was expected to be the highest, reflecting the learning from repeated similar tasks plus the effect of the tool on performance. Some interaction was expected between the treatments, reflecting the tool effect and the learning effect between pre-test and post-test. In order to test these premises, three hypotheses were formed and expressed in the conventional null hypothesis and alternate hypothesis sets.

The first hypothesis set reflected the premise that an interaction exists between the mean treatment effects and the mean learning effects. These hypotheses are stated:

H1₀: There will be no interactions between the treatments and the tests.

H1₁: The post-test mean score will be higher than the pre-test mean score, and the tool treatment mean post-test score will be higher than the control treatment mean post-test score.

The rejection of the null hypothesis would indicate the presence of an interaction between the test and treatment factor effects.

The second hypothesis pair concerns the difference between tests. If a difference exists between the pre-test and post-test results, it will reflect the test effects due to the uses of different cases for the pre-test and post-test. This hypothesis is based on the expectation of a positive learning effect. This pair of hypotheses are expressed as:

H2₀: There will be no difference between the pre-test mean score and the post-test mean score.

H2₁: The mean score on the post-test will be greater than the mean score on the pre-test.

Again, a significant difference between test effects would result in the rejection of the H2₀ hypothesis.

Finally, the third pair of hypotheses addressed a difference between treatments (control versus tool). The tool was expected to enhance the tool group's performance. The tool group's mean score for the post-test was expected to be greater than the post-test mean score for the control group.

H3₀: The tool group's post-test mean score will be less than or equal to the control group's post-test mean score.

H3₁: The tool group's post-test mean score will be greater than the control group's post-test mean score.

If the experimental results indicate a significant difference between the treatment effects, then the H3₀ hypothesis must be rejected.

5. USER SURVEY

In addition to the quantitative experiment, a measure of the user analyst's perception of the value or utility of the CAASS was sought. User feedback on the content, accuracy, ease of use and format of the CAASS system was desired. No comprehensive instrument for computing success was discovered; however the Doll and Torkzadeh End-user Computing Satisfaction instrument presented accepted measures of user satisfaction (McHaney 1998). Other user surveys focus on all systems and services of an information systems department while the Doll and Torkzadeh instrument focused on the individual application (Goodhue 1998). Doll and Torkzadeh (1988) contended that end-user satisfaction is a surrogate for utility and can be used to measure the satisfaction or utility that the CAASS system provides.

Doll and Torkzadeh used 12 closed-ended and three open-ended questions in their questionnaire. The close-ended questions were evaluated by using a five-point Likert-type response scale, which helped minimize the fatigue of the respondent (Kendall 1999). The open-ended questions were used as global measures of overall user satisfaction (Figure 4).

The CAASS user survey is a modification of Doll and Torkzadeh's pre-tested and validated instrument. In order to reduce the time needed by the subjects to complete the survey, ten closed-ended questions and two open-ended questions were included in the CAASS instrument. The first closed-ended question was original. The remaining closed-ended questions were selected from the questions posed by Doll and Torkzadeh. The questions were designed to provide user feedback on four aspects of the CAASS system. The aspects are content, accuracy, format and ease of use. Closed-ended questions one, two and ten measure the user satisfaction with the format of the CAASS. Questions six and nine were focused on the user assessment of the tool's accuracy. The fourth, and fifth questions addressed the users' impressions about CAASS content and questions three, seven and eight focused on the tool's ease of use.

The first open-ended question was included to verify that the users had completed the last three instruction steps while using the tool. The second open-ended question was from the Doll and Torkzadeh instrument and served as a global measure of satisfaction.

6. EXPERIMENTAL RESULTS

Analysis of Variance (ANOVA)

The experimental results were analyzed initially by using an ANOVA with repeated measures over the test factor. There was a significant interaction that caused rejection of the no interaction hypothesis H_{10} . This rejection was based on the treatment by test interaction F-test result: $F(1,70) = 14.54$; $p = 0.0003$.

Since an interaction occurred, H2 and H3 each must be decomposed and each part evaluated in order to understand the interaction effect. Hypothesis H2 reflected the expectation that the control group post-test scores would show a positive learning effect. Additionally, the post-test scores for the tool group was expected to show a learning effect plus a positive tool effect. Hypothesis H2 was separated into H2a and H2b to investigate these effects.

H2a₀: There will be no difference between the control group's pre-test mean score and the control group's post-test mean score.

H2a₁: The control group's post-test mean score will be greater than the control group's pre-test mean score.

The H2a₀ hypothesis was rejected based on a paired student's t test result of $t(0.05, 70) = 4.159388$; $p = 0.0001$.

As expected, the control group's post-test mean score of 48.35 was a significant improvement over the control group's pre-test mean score of 35.89. This increase in the control group's mean score reflected a positive learning effect. The control group's repetition of similar tasks by performing the pre-test and post-test, had produced a rise in performance level.

Hypothesis H2b was expressed as:

H2b₀: The tool pre-test mean score will be less than or equal to the tool post-test mean score.

H2b₁: The tool post-test mean score will be greater than the tool pre-test mean score.

The H2b₀ hypothesis could not be rejected with $t(0.05, 70) = -1.23382$; $p = 0.7786$.

The failure to reject the H2b₀ premise indicated that there was no significant increase in mean score between the tool group's pre-test mean score of 36.52 and the post-test mean score of 32.82. In fact, a decrease in the mean test score was observed. The expected post-test rise in the tool group's mean score, reflecting the positive learning effect from repeating similar tasks and the additional positive effect from tool use, was not realized. The slight decrease in the tool group's post-test mean score possibly reflects the negative impact of the lack of hands-on training in tool use. The use of the tool apparently nullified the positive learning effect from repeating similar tasks.

Hypothesis H3 became the hypotheses pair H3a and H3b. Hypothesis, H3a, serves to further investigate tool versus control on the post-test. H3b provides a consistency check on the pre-test performance of the tool and control groups.

CAASS USER SURVEY

This questionnaire is an evaluation instrument for the Case-based Adaptive Analyst Support System (CAASS). In the questions below CAASS will be referred to as “the system” or “the application”.

Please check one box for each of the ten questions below.

1. Do you prefer the system to the manual method?
 Almost never Some of the time About half of the time Most of the time Almost always
2. Do you think the output is presented in a useable format?
 Almost never Some of the time About half of the time Most of the time Almost always
3. Is the system difficult to operate?
 Almost never Some of the time About half of the time Most of the time Almost always
4. Does the system provide sufficient information?
 Almost never Some of the time About half of the time Most of the time Almost always
5. Do you find the output relevant?
 Almost never Some of the time About half of the time Most of the time Almost always
6. Is the system successful?
 Almost never Some of the time About half of the time Most of the time Almost always
7. Is the system easy to use?
 Almost never Some of the time About half of the time Most of the time Almost always
8. Is the system user friendly?
 Almost never Some of the time About half of the time Most of the time Almost always
9. Do you think the system is reliable?
 Almost never Some of the time About half of the time Most of the time Almost always
10. Overall, how would you rate your satisfaction with this application?
 Nonexistent Poor Fair Good Excellent
11. What evidence did you see that would indicate the system adapted or learned with repeated use?

12. What aspects of the application, if any, were you most satisfied with and why?

Figure 3. CAASS User Survey Questionnaire.

H3a₀: The tool group’s mean score on the post-test will be less than or equal to the control group’s mean score on the post-test.

H3a₁: The tool group’s post-test mean score will be greater than the control group’s post-test mean score.

The H3a₀ hypothesis cannot be rejected based on $t(0.05,70) = -5.18185$; $p = 0.9999$.

Hypothesis H3a tested a comparison between the tool group’s post-test performance and the control group’s baseline performance. The failure to reject H3a₀ implies

that the tool group’s post-test performance was less than the baseline mean scores established by the control group. The tool group’s post-test mean score of 32.82 was considerably below the control group’s baseline mean score of 48.35. This result was consistent with the failure to reject the hypothesis of no increase in the tool group’s post-test mean score.

The H3b hypotheses pair are stated:

H3b₀: The tool group’s pre-test mean score will be equal to the control group’s pre-test mean score.

H3b₁: The tool group's pre-test mean score will not equal the control group's pre-test mean score.

H3b₀ cannot be rejected based on a t-score of :
 $t(0.05,70) = 0.211353$; $p = 0.8332$.

The comparison of the tool group's pre-test and the control group's pre-test mean scores was expected to indicate any significant difference in skill level. The failure to reject H3b₀ indicates that no significant difference was detected between the control group's and tool group's pre-test mean scores. The tool group's pre-test mean score of 36.52 and the control group's pre-test mean score of 35.89 demonstrate a similar skill level on

the pre-test manual task. The tool group's mean score was not substantially different from the control groups established baseline mean score.

User Survey Results

All 36 of the CAASS users in the tool group responded to the closed-ended questions. The CAASS survey questions and the modal response for each question is shown in Table 2 below. Each response was based on the Likert scale choices shown in Figure 3. The survey responses support the overall design of the tool with respect to format, content, accuracy and ease of use. The survey results indicate an overall satisfaction with the tool in spite of the lower experimental performance.

Table 2. Responses to CAASS End-User Survey

Question No.	Question	Modal Response	Number	Mean	Std. Dev.
CAASS Format					
1.	Do you prefer the system to the manual method?	Almost Always	18 of 36	4.12	1.004
2.	Do you think the output is presented in a useable format?	Most of the Time	22 of 36	3.56	0.896
10.	Overall, how would you rate your satisfaction with this application?	Good	26 of 36	3.78	0.790
CAASS Content					
4.	Does the system provide sufficient information?	Most of the Time	20 of 36	3.49	0.840
5.	Do you find the output relevant?	Most of the Time	20 of 36	3.34	0.938
Accuracy					
6.	Is the system successful?	Most of the Time	27 of 36	3.83	0.781
9.	Do you think the system is reliable?	Most of the Time	23 of 36	3.80	0.872
Easy to Use					
3.	Is the system difficult to use?	Some of the Time	22 of 36	1.927	0.905
7.	Is the system easy to use?	Most of the Time	20 of 36	3.88	0.899
8.	Is the system user friendly?	Most of the Time	20 of 36	3.83	1.070

Only 50% of the students completing the survey chose to answer open-ended question 11 concerning the ability of the tool to learn or adapt with repeated use. Ten subjects indicated that the system did learn and provided the learned case as the best match with repeated use. Eight students stated they observed no evidence of learning. Since all eighteen of the students responding to this question had complete templates recorded for the post-test, a possible explanation may be that the eight students failed to carry out the last three steps in the written instructions.

These steps included allowing the system to learn the current organization parameters and conduct a second trial to see if the tool would find the learned results as the best match.

Twenty-six students, or 72%, chose to answer the second open-ended question (question 12) concerning aspects of the tool with which they were most satisfied. Twenty-three of the students generated responses that could be grouped into the three general positive responses and three students generated the dislike responses that follow (Table 3 below).

Table 3. Question 12 responses.

LIKES	DISLIKES
Screen sequences and format (visual prompts, pull-down menus and combination boxes)	Needs more detailed help screens
Ease of use of the tool helped complete the fact gathering tasks	Didn't like the application at all. I had not idea what I was doing
Best fit case match helped complete current case fact gathering	No advantage over manual method

The help function definitely needs expansion. Although this was a CAASS system, an expanded help may have reduced the impact of the lack of hands-on training. Hands-on training may have helped the student who responded "I had no idea what I was doing."

The response that the tool offered no advantage over the manual method was probably accurate for the small case studies that allowed completion within the 90-minute time allotted. Use in the fact gathering activities in a real world organization would have provided a better assessment of the tool's possible advantage.

7. CONCLUSIONS

The tool group's manual performance was measured and compared to the control group baseline. Both groups performed as expected. There was no significant pre-test difference between the groups using manual

methods. This established that both groups possessed similar skill sets and performed similarly on the same task. The tool group was given a set of detailed operating instructions and allowed to use the tool while completing the post-test task. It was expected that both control and tool groups would benefit from the learning effect from repeating similar tasks. Additionally, the tool group was expected to realize an increase in performance above the control baseline due to the positive contribution of the tool. This did not occur. The tool group's performance apparently suffered from a lack of hands-on training in the use of the tool.

Familiarity with the Windows environment presented by the tool's Visual BASIC user interface may have caused a degree of automated task completion by the user. Users may have selected the first likely alternative suggested by the tool without performing a cognitive evaluation as to which suggested alternative best fit the task under study.

The experimental results did not demonstrate a tool performance advantage over the manual method for the pre-post test tasks of limited scope and duration. More complex real-world tasks may better demonstrate the tool's capabilities.

The users' utility for the tool was relatively high. The results of the CAASS user survey provided some insight into the level of user utility for the CAASS in the areas of tool format, content, accuracy, and ease of use. The users' preferred the use of the tool to the manual method. They felt the tool's output was in a useable format and were satisfied with the overall application format.

The CAASS content was believed to provide sufficient relevant information to complete the tasks assigned. The exception to this view was the CAASS help facility. Users would like to see a more extensive development of the help function. This is a valid observation. The CAASS provided a "bare bones" help facility for major tool functions. A production tool will require a greatly expanded help facility.

The users' overall evaluation of the CAASS was good. The users felt the CAASS was reliable and produced successful results. The familiar Windows interface was adjudged user friendly and easy-to-use. As mentioned earlier, this familiar easy to use interface, in the absence of hands-on-training may have contributed to the unexpected below baseline performance of the tool group on the post-test task. A software tool the users have not yet made an automatic process may be more hindrance than help. Despite the higher level of familiarity students have with a Microsoft Windows-like interface, the GUI is not enough to generate automatic processing and avoid performance degradation.

8. REFERENCES

- Collins, Rosann Webb, 1993, Impact of Information Technology on the Processes and Performance of Knowledge Workers. Doctoral Thesis, Graduate School, University of Minnesota.
- Cook, Thomas D. and Donald T. Campbell, 1979, Quasi-Experimentation. Houghton Mifflin, Boston, MA.
- Dennis, Alan, and Barbara Haley Wixom, 2000, Systems Analysis and Design. John Wiley & Sons, New York, NY.
- Doll, William J., and Gholamreza Torkzadeh, "The Measurement of End-User Computing Satisfaction." MIS Quarterly, 1988, 12.2, pp. 259-274.
- Easton, Annette, and George Easton, 1996, Cases for Modern Systems Analysis and Design. Benjamin/Cummings, Menlo Park, CA.
- Goodhue, Dale L. "Development and Measurement Validity of a Task-Technology Fit Instrument for User Evaluations of Information Systems." Decision Sciences, 1998, 29.1, pp. 105-138.
- Haag, Stephen, Maeve Cummings, and James Dawkins, 2000, Management Information Systems for the Information Age. 2nd Ed. Irwin McGraw-Hill, Boston, MA.
- Hoffer, Jeffery A., Joey F. George, and Joseph S. Valacich, 1999, Modern Systems Analysis and Design. 2nd Ed. Addison-Wesley, Reading, MA.
- Keppel, Geoffrey. 1991, Design and Analysis. Prentice Hall, Englewood Cliffs, NJ.
- Kendall, Kenneth E. and Julie E. Kendall, 1999, Systems Analysis and Design. 4th Ed. Prentice Hall, Upper Saddle River, NJ.
- McHaney, Roger, and Timothy P. Cronan, "Computer Simulation Success: On the Use of the End-User Computing Satisfaction Instrument: A Comment." Decision Sciences 29.2, 1998, pp. 525-536.
- Montgomery, Douglas C. 1991, Design and Analysis of Experiments. 3rd ed. John Wiley & Sons, New York.
- Satzinger, John W., Robert B. Jackson, and Stephen D. Burd. 2000, Systems Analysis and Design in a Changing World. Course Technology, Cambridge, MA.
- Schneider, Walter and Arthur D. Fisk, "Degree of Consistent Training: Improvements in Search Performance and Automatic Process Development." Perception & Psychophysics, Vol. 31, No. 2, February 1982, pp. 160-168.
- Thorngate, Warren. "Must We Always Think Before We Act?" Personality and Social Psychology Bulletin, Vol. 2, No. 1, Winter 1976, pp. 31-35.