A Psychometric Assessment of Measures of Scripts in Consumer Memory

RUTH ANN SMITH
MICHAEL J. HOUSTON*

This study examines the psychometric qualities of two measures of scripts for a consumer event. A laboratory experiment in which the measures' reliability and validity were evaluated produced some questions about the stability of retrospective self-report measures. However, the results were generally encouraging for the psychometric properties of an alternative measure employing a recognition task. We suggest further development of this measure to facilitate large-scale investigations of script-based processing in the consumer context.

There is a large body of research on the structure and function of memory that documents memory's influence on human behavior in general and on consumer behavior in particular. One conclusion that derives from this literature is that no single explanation of memory is entirely adequate to account for the diversity of human information processing strategies. Rather, both the nature of the to-be-stored information and expectations about its future use apparently influence memory structure.

The present research involves one particular type of memory structure—scripts—that has been the subject of increasing attention in the recent consumer-behavior literature. However, both in the consumer context and elsewhere, the majority of script research has relied on measurement procedures of doubtful merit. Therefore, the present investigation seeks to evaluate the psychometric status of two script measures in an effort to facilitate further research of script-based processing in the consumer setting.

The following discussion reviews the nature and function of scripts and summarizes relevant empirical findings. Next, it describes the methodological tradition that has prevailed in script research. Finally, it advances several propositions concerning the reliability and validity of two script measures and considers an empirical test of those propositions.

SCRIPTS

Scripts are one type of the broad classification of memory structures generally known as schemata. A schema contains an individual's general knowledge about some stimulus domain, represented in generic rather than episodic form. A schema is not an exact record of a specific stimulus. Rather, schemata contain abstractions teased out of environmental regularity (Mandler 1982; Rumelhart and Ortony 1977). There are several types of schemata, and they are categorized according to the stimulus domain that acts as their organizing theme.

A script is a type of schema for which the organizing theme is a repetitive event. As defined by Schank and Abelson, a script is "a predetermined, stereotyped sequence of actions that define a well-known situation" (1977, p. 41). Scripts are distinguished from other types of schemata by two definitive qualities: first, a script contains a set of component actions, and second, those actions are related in a causal temporal sequence. Examples might be dining in a restaurant, visiting a physician, or making a bank deposit.

Scripts appear to develop a hierarchical structure in memory as an individual experiences multiple occurrences of highly programmed events (Martin, Harrod, and Siehl 1980; Nelson 1981). Once formed, a script's general function is to facilitate cognitive processing. More specifically, when an incoming stimulus activates a script, a rich network of information is tapped that substantially reduces the burden of processing the stimulus. Scripts appear to exert significant influence on both encoding and retrieval, with the result that script-based processing is highly automatic and efficient, if not always accurate.

During encoding, a script acts as a framework for selecting and organizing relevant information. For example, Bower, Black, and Turner (1979) reported that information central to an event is more likely to be recalled than information that is irrelevant. In addition, Bransford and Johnson (1972) found that

*Ruth Ann Smith is Assistant Professor of Marketing, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. Michael J. Houston is McManus-Bascom Professor of Marketing, Graduate School of Business, University of Wisconsin, Madison, WI 53706. The authors gratefully acknowledge the contributions of Arthur Glenburg to the development of this research.

recall of a group of related actions was superior if an organizing theme that activated a script was provided during encoding. Empirical support for a script's influence on retrieval is also available. Several studies have demonstrated both intrusion errors during recall and false recognition of script-consistent, but missing, actions (Bower et al. 1979; Graesser, Gordon, and Sawyer 1979; Graesser et al. 1980). These retrieval biases suggest that when a stimulus configuration is incomplete, the missing information can be inferred from the generic script for an event.

Recognition of the relevance of scripts to consumer behavior is fairly recent, but the limited research available in this context supports both the presence of scripts for consumer events and their influence on consumer information processing. For example, culturally uniform scripts for dining in a restaurant (Bower et al. 1979; John and Whitney 1982) and shopping (Whitney and John 1983) have been reported. Furthermore, Whitney and John (1983) provide evidence that scripts influence the encoding and retrieval of information about consumers' experiences. It should be noted, however, that the status of schema theories in general, and of scripts in particular, is controversial. In an extensive review of schema research, Alba and Hasher (1983) argue that many studies, including some of those cited above, are subject to alternative explanations. This problem may derive in part from inadequacies in both theory specification and measurement procedures. The latter is the primary concern of the present investigation.

In examining the methods used to quantify scripts, it is evident that the research to date has relied almost exclusively on retrospective self-report measures. Table 1 summarizes a sample of script investigations and reveals two common purposes for which such measures have been used. The first is to determine a script's constituent structure (Bower et al. 1979; Graesser et al. 1979). The second is to use this script as the stimulus for subjects' long-term memory reconstructions. These reconstructions constitute the dependent measure that is used to evaluate a script's influence on various processing stages (Alba et al. 1981; Bower et al. 1979; Bransford and Johnson 1972; Lichtenstein and Brewer 1980) or to analyze the script's organizational structure (Abbott and Black 1980; Black and Bern 1981; Martin et al. 1980).

This pervasive reliance on retrospective self-reports is troublesome in two regards. First, it has been argued that such measures may represent only a respondent's inferential processes—not the contents of memory (Ericsson and Simon 1980; Nisbett and Wilson 1977). Second, virtually the only attempt to ascertain the psychometric qualities of such measures in the script context has been to examine interjudge reliability. And as revealed in Table 1, few of the reviewed investigations addressed even this issue. Moreover, the procedures employed to code responses in these studies are often imprecisely described or potentially subjective in the extreme. Thus, a very weak basis exists for determining reliability, and virtually no evidence is available upon which to establish validity.

The present research was intended to specifically address the reliability and validity of scripts in the consumer context. Although the magnitude of scripts' influence on consumer behavior has not been clearly demonstrated, the few studies to date suggest that script-based processing does occur in certain consumer settings (Bower et al. 1979; John and Whitney 1982; Whitney and John 1983). While the growing recognition that such automatic—perhaps even mindless—processing strategies are prevalent in consumer decision making (Kassarjian 1978; Olshavsky and Granbois 1979) would certainly justify further script research in consumer settings, such efforts must necessarily depend on reliable and valid measurement procedures. This study is an initial step toward that goal.

PROPOSITIONS

As noted previously, efforts to establish the psychometric qualities of retrospective self-report measures of scripts have been restricted to determining the level of agreement among coders. The two studies included in Table 1 that confronted this reliability issue reported impressive results, with 95 percent (Alba et al. 1981) and 99 percent agreement (Martin et al. 1980). However, the majority of the investigations reviewed failed to indicate either the level of reliability obtained or the nature of the coding procedures employed. Development of a formalized coding technique that minimizes subjectivity is a requisite to determining the equivalence of judges for such measures.

In addition, evidence of the test-retest reliability of self-report measures is needed. Although several theorists agree that scripts may be modified as new experiences occur (Abelson 1976; Bartlett 1932; Rumelhart and Ortony 1977; Schank and Abelson 1977), it is reasonable to assume that scripts must be somewhat stable through time to facilitate information processing. Consequently, stability would also be expected in the responses of individuals with scripts to measures of such memory structures. Moreover, since individuals without a script for an event would presumably have to rely on some constructive reasoning process when responding to a script measure, less stability would be expected in these responses than in those of individuals who rely on a script.

A demonstration of the validity of retrospective self-report measures is also necessary. Churchill states (1979, p. 70):

> A fundamental principle in science is that any particular construct or trait should be measurable by at least two,

TABLE 1

UTILIZATION OF RETROSPECTIVE SELF-REPORTS IN SELECTED SCRIPT INVESTIGATIONS

Author	Purpose of self-report	Coding procedure	Interjudge reliability
Abbott and Black (1980)	Unaided recall of two prose narratives to determine the cognitive structure of scripts in memory	Not specified	Not specified
Alba, Alexander, Hasher, and Caniglia (1981; Experiment 1)	Unaided recall of a prose narrative to determine the influence of context on script-based processing	Same as Bransford and Johnson (1972)	95%
Black and Bern (1981; Experiment 2)	Unaided recall of four prose narratives to evaluate the influence of causal and temporal relations of stored script actions	"Loose gist" criterion used to match responses to target sentences in the narratives	Not specified
Bransford and Johnson (1972)	Unaided recall of a prose narrative to determine the effect of script activation on understanding	Two judges compared responses to a priori-designated "idea units;" disagreements resolved by a third judge	Not specified
Bower, Black, and Turner (1979)			
Experiment 1	Unaided recall of a commonplace event to determine the cultural uniformity of scripts	Not specified	Not specified
Experiment 3	Unaided recall of eighteen prose narratives (derived from the responses in Experiment 1) to evaluate the influence of scripts on retrieval	Not specified	Not specified
Experiment 7	Unaided recall of six prose narratives to determine memory for script deviations	Not specified	Not specified
Graesser, Gordon, and Sawyer (1979; Experiment 1)	Unaided recall of eighteen commonplace events to develop stimulus materials for subsequent experiment	Not specified	Not specified
Graesser, Woll, Kowalski, and Smith (1980; Experiment 2)	Free recall of actions that typically occur in commonplace situations, or free recall of one of eight prose narratives to determine the influence of a script on retrieval of typical and atypical actions over varying retention intervals	"Gist" criterion	Not specified
Lichtenstein and Brewer (1980; Experiments 2–5)	Unaided recall of a videotaped scene or prose narrative to evaluate the influence of scripts on retrieval	"Loose gist" criterion used in comparing responses to action units identified in a norming study	Not specified
Martin, Harrod, and Siehl (1980)	Unaided recall of one to four prose narratives to determine the effect of repetition and similarity on script acquisition	Single-coder content analyzed responses according to three distinct sets of clearly specified criteria	99% agreement between coder and second judge on a 10% random sample of responses

> and preferably more, different methods. Otherwise the researcher has no way of knowing whether the trait is anything but an artifact of the measurement procedure.

Therefore, if self-reports represent the contents of a script, it would be expected that subjects' scores would converge with the scores for other reliable script measures.

However, the literature offers a notable lack of alternatives to self-reports which, in addition to their unclear psychometric status, also pose certain other problems for script investigations. First, the cognitive demands of generating a response are substantial and seem incompatible with the automatic nature of script-based processing. Moreover, in the consumer context, the relevance of scripts derives in part from their universality across large groups: executing investigations on the required scale is cumbersome at best, due to the elaborate coding procedures needed to analyze responses to such measures.

Consequently, an alternative script measure is needed that not only permits an assessment of the validity of self-reports, but that is also more compatible with the inherent character of script-based processing and facilitates large-scale studies of consumers' scripts. Such an instrument—the rank-order measure—was developed in this investigation. Respondents to this measure performed a recognition task in which they distinguished between actions relevant to and irrelevant to an event and arranged the former in the script-defined order. This task captured the two definitive qualities of a script; i.e., that it consists of a group of component actions, and that the actions relate in a causal temporal sequence. Contingent on a demonstration of test-retest reliability, responses to this measure would be expected to converge with those made to a self-report measure of scripts.

A final step in addressing the psychometric status of script measures is to assess construct validity through a comparison of responses to the two measures of individuals with and without a script. It is expected that those respondents with a script will have access to a rich network of information unavailable to those without such a memory structure. Therefore, the former group should outperform the latter on script measures as indicated by significantly higher scores. This discussion has provided a conceptual basis for the following propositions, in which the terms schematic and aschematic are used to denote individuals with and without a script for a particular event.

P1: Retrospective self-reports obtained from schematics and aschematics will exhibit interjudge reliability as indicated by a high level of agreement among independent coders.

P2: The responses of schematics to script measures obtained on two occasions will exhibit significantly more stability than those of aschematics.

P3: There will be a strong positive correlation between the relative frequencies with which component actions of a script are included in schematics' responses to different script measures.

P4: The scores of schematics on both retrospective self-report and rank-order measures of scripts will be significantly higher than those of aschematics.

METHOD

Context

The propositions were investigated in the context of the placement service operated by a large midwestern university's School of Business. The placement-service setting was appropriate for several reasons. First, the highly structured and repetitive character of the placement-service event would facilitate script formation among its clients. Second, although both actual and potential clients had various qualities in common (academic skills, professional aspirations), actual clients had accumulated extensive service-event experience, while potential clients had quite limited experience. It was therefore possible to identify schematics and aschematics whose other between-group differences were minimal. Finally, the placement-service event shares the conceptual features of many consumer services: it is both impermanent and intangible and it requires the joint participation of a consumer/client and service provider. Moreover, though a direct exchange process does not occur in this service, such an atypical exchange is not uncharacteristic of various other consumer services such as shopping or banking. Since the unique exchange process does not alter clients' use of such services relative to those involving a direct monetary transaction, the measurement of related scripts can reasonably be expected to be representative of many service events.

Design

The propositions were evaluated in a laboratory experiment that utilized a 2 × 2 factorial design. The first factor—subject type—included both schematics and aschematics. Subjects were categorized using their responses to a 39-item questionnaire that covered level of experience with and general knowledge of the service event. Inasmuch as knowledge structures for events develop through repetitive experience (Martin et al. 1980; Nelson 1981), responses to the questionnaire were expected to be related to the presence or absence of a schematic representation of the event.

The questionnaire was completed by 106 volunteer subjects: one half were actual, and the remaining half were potential clients. Cronbach's alpha was used to evaluate the internal homogeneity of the measure, which was found to be quite satisfactory at 0.97. The median score of the responses was 94.5 (possible range of 3 to 148). Subjects scoring below this value were categorized as aschematic, while those whose scores exceeded the median were classified as schematic. From each group, 40 subjects were randomly selected to participate in the study.

The second factor in the design was a measurement condition. Twenty subjects of each type were randomly assigned to respond to either a retrospective self-report or a rank-order measure. The former was modeled after similar measures used in previous script research (Bower et al. 1979; Graesser et al. 1979; John and Whitney 1982; Whitney and John 1983), and utilized an open-ended probe requesting a written description of the placement-service event. The rank-order measure consisted of a randomly ordered list of 35 actions, only 20 of which were relevant to the focal event. Respondents were to identify the relevant actions and number them sequentially to indicate the normal order of occurrence.

Development of the Script Measures

The specific wording of the self-report measure was developed through extensive pretesting to provide subjects with clear indications of (1) where to begin and end their accounts, and (2) the appropriate level of detail. In developing the rank-order measure, it was necessary to identify the set of actions and the temporal sequence typical of the focal event. We accomplished this by using a procedure adapted from Bower et al. (1979) to analyze 10 pretest self-reports obtained from former clients of the service. These self-reports were edited to determine how frequently various actions were mentioned. Using a 30 percent relative frequency as a criterion level, 20 of 33 distinct actions were selected as the relevant actions for inclusion in the rank-order measure.

The script-defined order of these actions was then determined through a paired-comparison technique developed by John and Whitney (1982). The results presented in the Figure indicate a high level of agreement about the order of the 20 base actions. This agreement is evident from the small number of entries in the portion of the matrix below the main diagonal. The dominant order was therefore adopted as the typical sequence of the component actions.

Since there was concern about possible modifications to the focal event after the exposure of the pretest subjects, both the base actions and the sequence identified from the pretest procedure were subsequently verified by analyzing the experimental self-reports of schematic subjects. Not only did the same 20 constituent actions emerge, but the temporal order was

FIGURE

PAIRED COMPARISON OF SEQUENTIAL ORDER OF BASE ACTIONS

Action Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1			5	4	2	5	2		2		1	5	1	1	4	1	3	2	2	2
2			1	1		1		1			1	1		1						
3				7	3	7	3	1	1	1	2	7	1	2	6	1	5	4	3	2
4					3	6	2	1	1	1	2	6	1	2	5	1	4	3	3	2
5						1					1	2	1	1	2		1	1	1	1
6					1		3	1	2	1	2	7	1	2	6	1	5	4	3	2
7									1			2			3	1	3	2	2	
8											1	1		1						
9												1			2	1	1		1	
10												1			1		1	1		1
11												2		2	1		1	1	1	1
12						1			1				1	2	6	1	5	4	3	2
13																				
14															1		1	1	1	1
15																1	5	4	3	2
16																	1		1	
17																		4	3	2
18																			2	2
19																				1
20																				

NOTE: Cell entries should be interpreted as the number of respondents who listed the row action before the column action.

replicated. Therefore, these sequenced actions (reproduced in Appendix A) seem to adequately capture the constituent structure of the script.

A group of irrelevant actions was included in the rank-order measure in addition to the relevant actions. Each of these actions was generally appropriate to job hunting, but would not normally occur in the context of the focal event. We obtained these from a convenience sample of former clients who rated 31 potentially irrelevant actions on a five-point scale with anchors of "I never did this" (scale value = 1) and "I always did this" (scale value = 5). The 15 actions satisfying the criteria of a mean scale value ≤ 2.0, s.d. ≤ 1.5, and the relative frequency of a low rating (1 or 2) ≥ 60 percent were combined in random order with the 20 base actions to constitute the rank-order measure. These irrelevant actions are reproduced in Appendix B.

Procedure

The experiment consisted of two sessions separated by a one-month interval to minimize memory effects in the second set of responses. Subjects randomly assigned to either the rank-order or self-report measure were run in small groups, with all sessions conducted in a formal setting under the supervision of one of the investigators. At the first session, subjects completed their assigned measure, using as much time as desired. Six subjects failed to attend this session, resulting in a sample composed of 36 schematics and 38 aschematics.

At the second meeting, 33 schematics and 36 aschematics completed a questionnaire about their experience with the focal event after the first session. This measure was used to identify subjects who were highly likely to have a true change in the mental representation of the event. Although no aschematic reported having had any interviewing experiences, two schematics who reported "unusual" interviews were dropped from the analysis of the second set of responses. Subjects then responded to their assigned measure—rank-order or self-report—a second time, thus completing the tasks.

The procedure was identical for the two measurement conditions, with the exception that while a single version of the self-report measure was used for both sessions, two versions of the rank-order measure were used, each with a different order for the 35 actions. Using two versions of the rank-order measure was intended as an additional safeguard against memory effects.

Coding Procedures

In order to evaluate the several propositions, it was necessary to analyze subjects' responses so that both the presence and order of the base actions could be ascertained. In the rank-order condition, a base action was deemed present if it was among those selected as relevant to the focal event. In the self-report condition, a coding procedure was developed to accomplish this purpose. In devising this procedure, the primary objective was to minimize subjectivity insofar as possible. Three judges, all naive to the placement service event and the purpose of the investigation, were selected. Using the sequenced base actions as benchmarks, they were instructed to bracket any part of a self-report they felt was descriptive of one of the actions, and to indicate its identifying number. Each judge was trained in this task prior to coding experimental responses and each judge worked totally independently. The investigators' only involvement in this process was to tabulate the three judges' decisions. If at least two judges indicated similar origination and termination points and used the same action number, a base action was considered to be present in a response. No mechanism to resolve coder disagreements was used.

RESULTS

Reliability

Proposition 1 predicted a high level of interjudge agreement and was evaluated using a procedure developed by Holsti (1969):

$$R = \frac{N(\text{average interjudge agreement})}{1 + [(N - 1)(\text{average interjudge agreement})]}$$

where:

N = number of judges (3)

and average interjudge agreement is calculated as:

$$\frac{\frac{2M_{12}}{n_1 + n_2} + \frac{2M_{23}}{n_2 + n_3} + \frac{2M_{13}}{n_1 + n_3}}{N}$$

where:

N = number of judges (3)

M = number of coding decisions upon which there is agreement between any two judges

n_1, n_2, n_3 = number of coding decisions by judges 1, 2, and 3, respectively

This procedure revealed 95 percent and 96 percent agreement for schematics' self-reports at Sessions 1 and 2, respectively. For aschematics, 90 percent agreement was obtained for the first session responses and 92 percent for the second. These results indicate that the coding technique employed provided a reliable basis for additional analyses.

Proposition 2 was concerned with the test-retest reliability of the script measures, and predicted that the responses of schematics would exhibit significantly greater stability through time than would those of aschematics. To evaluate this proposition, two scores representing the configuration of responses were compared between the subject types. Configuration was defined as the degree of similarity between the action sequence in script-measure responses obtained on two occasions. The first configuration score—the duplicated action score—represented the percentage of the total number of base actions contained in a response pair that were duplicated. The feasible range of scores was from 0 to 100 percent for complete duplication of actions in the two responses. The second configuration score—the duplicated sequence score—represented stability in the order of the duplicated actions. A duplicated action was judged to be in the same sequential order in two responses if it occupied either the same absolute position (e.g., it was the i^{th} action) or the same relative position (e.g., it followed the i^{th} action). The feasible range of scores was again 0 to 100 percent, where 100 percent indicated total replication of the sequence of duplicated actions.

These mean configuration scores, which are displayed in Table 2, represent a very stringent test of test-retest reliability since they reflect much more than the degree to which scores correlate across administrations of the measures. Specifically, the configuration scores reveal the degree to which subjects generated

TABLE 2

CONFIGURATION OF SCHEMATIC AND ASCHEMATIC SUBJECTS' RESPONSES

Responses	Schematics	Aschematics	t	p
Self-report measure[a]				
% Duplicated actions	58.2	55.2	.51	.307
% Duplicated sequence	86.2	90.3	−.52	.305
Rank-order measure[b]				
% Duplicated actions	91.2	80.7	3.14	.002
% Duplicated sequence	83.1	69.3	3.59	.001

[a] For the self-report measure $n = 15$ for schematics and $n = 19$ for aschematics.
[b] For the rank-order measure $n = 16$ for schematics and $n = 17$ for aschematics.
NOTE: Scores are mean scores. Nine schematics and four aschematics were excluded from this analysis due to nonattendance at one or both experimental sessions or to a high probability of change in the mental representation of the event during the experiment.

exactly the same response on separate occasions. In the rank-order condition, as revealed in Table 2, it was found that schematics were significantly more prone than aschematics to duplicate base actions ($t = 3.14$, $p < 0.002$) and sequence ($t = 3.59$, $p < 0.001$). Although this finding supports the proposition, in an absolute sense aschematics' responses appear to be quite stable (80.7 percent duplicated actions; 69.3 percent duplicated sequence). Consequently, one might speculate that the actions composing the rank-order measure are so obvious in their fit (or lack of fit) to the focal event that the measure does not discriminate between schematics and aschematics.

However, several factors appear to argue against this interpretation. First, the proportion of aschematics who included any irrelevant actions in their responses was significantly greater than the proportion of schematics who included any irrelevant actions ($t = 9.95$, $p < 0.001$). Moreover, in both the first and second set of responses, aschematics included significantly more irrelevant actions ($t = 3.88$, $p < 0.001$; $t = 3.83$, $p < 0.001$) and significantly fewer relevant actions ($t = -3.59$, $p < 0.001$; $t = -2.30$, $p < 0.025$). Combined with the greater stability of schematics' responses, these results suggest that aschematics' configuration scores should not be interpreted to mean that the measure is overly simple. Not only are schematics more stable, but their responses differ from aschematics in the expected ways.

Table 2 also shows that in the self-report condition, the expected superiority of schematics was not observed for either duplicated actions ($t = 0.51$, $p > 0.10$) or sequence ($t = -0.52$, $p > 0.10$). This finding may be due to the combined effects of instability in schematics' self-reports (as indicated by only 58.2 percent duplicated actions), and unexpected stability in aschematics' responses (as indicated by a duplicated action score approaching that of schematics—55.2 percent). The inability of schematics to duplicate more actions across their responses may be indicative of the measure's incompatibility with script-based processing. That is, the free-recall task performed in the self-report condition may involve a much higher order of cognitive processing than the automatic, mindless strategy characteristic of script-based processing. If so, the appropriateness of self-reports as measures of scripts would seem subject to question. Alternatively, the simpler recognition task of the rank-order measure may be more compatible with script-based processing. This interpretation is supported by schematics' significantly higher duplicated action score on the rank-order measure (91.2 percent) compared to 58.2 percent on the self-report measure ($t = 6.56$, $p < 0.001$). The differential processing demands imposed by the recall versus recognition tasks do not seem to affect the order of the actions, however, since no significant difference in schematics' duplicated sequence scores was observed ($t = 0.44$, $p > 0.10$).

The unexpected stability of aschematics' self-reports may be related to two factors. First, in the absence of a script, these subjects would necessarily have relied on an elaborative processing strategy in generating an initial response. The second administration of the measure would have provided a cue to this process. Although some cueing effect was anticipated, the difficulty of generating a self-report in the absence of a script may have produced a cue of such strength that the one-month interval was inadequate compensation. Second, aschematics' self-reports were quite brief, with an average of only 5.4 coded actions compared to 8.9 for schematics. As the absolute number of actions declines, the probability of duplication due to chance increases. Combined with an elaborative processing strategy, this probability of duplication is likely to have inflated the configuration scores of aschematics. Regardless, it must be concluded that Proposition 2 is only partially supported, and that further evidence of the stability of the self-report measure is needed. Moreover, this finding establishes a potential limit on the validity of this measure.

Validity

This investigation's second area of concern involved evaluating the script measures' validity. Proposition 3 concerns convergent validity and predicts a strong positive correlation between the relative frequencies with which various base actions were included in schematics' responses to either script measure. A more direct test of this proposition—comparing scores on the two measures—would not be particularly meaningful because of the greater cognitive demands asso-

TABLE 3

RELATIVE FREQUENCY OF BASE ACTIONS IN SCHEMATIC SUBJECTS' RESPONSES

	Session I				Session II			
	Self-report		Rank-order		Self-report		Rank-order	
Action	*n*	%	*n*	%	*n*	%	*n*	%
1	10	58.8	19	100.0	11	73.3	16	100.0
2	2	11.8	12	63.2	0	0.0	12	75.0
3	7	41.2	19	100.0	2	13.3	16	100.0
4	17	100.0	19	100.0	15	100.0	15	93.8
5	14	82.3	18	94.7	12	80.0	15	93.8
6	14	82.3	19	100.0	14	93.3	16	100.0
7	8	47.1	19	100.0	9	60.0	16	100.0
8	2	11.8	19	100.0	1	6.7	14	87.5
9	3	17.6	17	89.5	3	20.0	16	100.0
10	1	5.9	13	68.4	0	0.0	11	68.8
11	3	17.6	18	94.7	7	46.7	15	93.8
12	17	100.0	19	100.0	12	80.0	16	100.0
13	7	41.2	19	100.0	10	66.7	16	100.0
14	9	52.9	19	100.0	4	26.7	16	100.0
15	6	35.3	19	100.0	5	33.3	15	93.8
16	7	41.2	19	100.0	8	53.3	16	100.0
17	5	29.4	19	100.0	5	33.3	16	100.0
18	10	58.8	10	52.6	8	53.3	9	56.3
19	3	17.6	19	100.0	7	46.7	14	87.5
20	3	17.6	13	68.4	1	6.7	13	81.3

ciated with the self-report measure's recall task. It would be expected, however, that those base actions most frequently selected on the rank-order measure would also be recalled most often in self-report responses. Consequently, the relative frequency with which the 20 base actions were mentioned provides an indirect index of convergence between the measures.

Table 3 provides a summary of these frequencies. The correlation between those observed in the self-report and rank-order responses obtained at Session 1 was 0.32 ($p < 0.09$). For the second session responses, the frequencies correlated at 0.38 ($p < 0.05$). While the magnitude of these correlations is modest, it should be noted that they reflect responses from different subjects in the two measurement conditions. Consequently, these results may be interpreted as encouraging (although not conclusive) evidence of convergent validity.

Proposition 4 addressed the construct validity of scripts and predicted that schematics would consistently outperform aschematics on script measures. To evaluate this proposition, three scores were developed as performance indicators. Each score represented the degree to which a response corresponded to the constituent structure of the script for the focal event, as represented by the sequenced base actions. Higher scores indicated greater correspondence to this structure.

The first scoring procedure assigned one point for each base action present and one point for every correctly ordered action pair in a response. Score 1 was the sum of all points, and could vary between 0 and 39 (20 base actions + 19 pairs). Score 2 included an adjustment for the omission of base actions and was computed as: score 1 − 0.5 (number of missing base actions).[1] The rationale for this correction is that a script should facilitate retrieval of the entire action set defining an event. Therefore, this correction improves the sensitivity of the score as an indicator of the agreement between a response and the script's constituent structure. The range of feasible scores using this procedure was −10 to 39.

This rationale is also relevant to the final scoring procedure, since a script should facilitate not only retrieval of all component actions, but also the differentiation of irrelevancies. Thus, score 3 adjusted for

[1] In establishing the magnitude of the correction for the omission of base actions or for the inclusion of irrelevant actions, it was assumed that the recall/recognition of every script action was independent of the recall/recognition of other actions. Therefore, there would be an equal (or 0.5) probability that any action would or would not be recalled/recognized. More realistically, this probability is conditional on the recall/recognition of other script actions. Consequently, the 0.5 correction is somewhat arbitrary, but was selected due to the complexity of calculating the more accurate conditional probabilities.

TABLE 4

COMPARISON OF SCHEMATIC AND ASCHEMATIC SUBJECTS' SCORES ON TWO SCRIPT MEASURES

Scores[a]	Schematics	Aschematics	t	p
Self-report measure[b]				
Score 1-1	15.53	9.95	3.42	.001
Score 1-2	9.77	2.79	3.41	.001
Score 2-1	16.40	8.74	5.11	.001
Score 2-2	10.87	1.24	5.07	.001
Rank-order measure[c]				
Score 1-1	32.4	27.7	4.56	.001
Score 1-2	31.5	25.6	4.49	.001
Score 1-3	31.5	24.5	5.12	.001
Score 2-1	33.1	28.3	4.38	.001
Score 2-2	32.2	26.2	4.33	.001
Score 2-3	31.9	25.1	4.35	.001

[a] For scores, the left hand digit denotes occasion 1 or 2, and the right hand digit denotes procedure 1, 2, or 3.
[b] For the self-report measure n = 15 for schematics and n = 19 for aschematics.
[c] For the rank-order measure n = 16 for schematics and n = 17 for aschematics.
NOTE: Scores are mean scores.

both the omission of base actions and the inclusion of irrelevant actions: score 3 = score 1 − 0.5 (number of missing base actions + number of irrelevant actions). Using this procedure, scores could vary from −17.5 to 39. Score 3 was only computed for the rank-order condition, since irrelevant actions were not coded in self-report responses. Comparing the mean scores for schematics and aschematics on each of the three procedures produced the results presented in Table 4. Regardless of the procedure used to quantify performance on a script measure, schematics were consistently superior to aschematics at $p < 0.001$. These results offer very strong support for Proposition 4.

DISCUSSION

The findings detailed in this paper provide mixed evidence for the psychometric status of self-report measures for scripts, but enhance the status of the rank-order measure as an alternative. Both the equivalence of judges and the stability of responses were evaluated to determine the reliability of the two measures. Although interjudge agreement—90 percent or above—was quite satisfactory, analyzing test-retest reliability raised some doubts about the compatibility of self-report measures with script-based processing. Responses to the rank-order measure, however, were found to be quite stable across two administrations. In addition to reliability, convergent and construct validity were evaluated. A modest level of convergence between the two measures was observed. Moreover, each instrument successfully differentiated known groups of schematics and aschematics, indicating construct validity.

Two major conclusions may be derived from these results. First, the inconclusive evidence for the stability of the particular self-report measure employed herein suggests that more caution should be exercised in using such measures for script investigations. The popularity of similar measures in previous script research remains problematic, and further clarification of the test-retest reliability of such measures is necessary.

Second, this investigation offers preliminary evidence for the reliability and validity of an alternative measure of consumers' scripts. Although further investigation of the rank-order measure is desirable, the findings reported here are very encouraging since this instrument offers a far more convenient measurement alternative. This is of particular importance in the consumer context because of the need to evaluate script-based processing in large groups. Further development of the rank-order measure may reduce many of the practical problems associated with such large-scale investigations. The relevance to consumer behavior of automatic, simplified information processing strategies like script-based processing would certainly justify such efforts.

[*Received July 1984. Revised December 1984.*]

APPENDIX A

Sequenced Base Actions

1. From the master list of firms interviewing, select those compatible with your career goals and qualifications.
2. Obtain the appropriate computer (Interview Request) forms at the Placement Office.
3. Prepare an Interview Request Form, a College Interview Form, and any other materials such as a coursework summary or a resume.
4. Turn all forms in to the Placement Office 7 days in advance of the firm's visit.
5. Check the firm's interview schedule to see if you have an appointment.
6. If on the firm's interview schedule, research the company using materials available in the Placement Office and from other sources.
7. Prepare a list of questions for the interviewer, and formulate and review answers to anticipated questions.
8. Purchase a good business suit.
9. Prepare materials to be taken to the interview.

10. Attend the pre-interview reception.
11. Dress appropriately.
12. Arrive 10–15 minutes before the scheduled time of the interview and check "Today's Interviews" for the location of the interview, and the interviewer's name, title, and address.
13. Wait outside the interview room to be called.
14. When called by the interviewer, introduce yourself, shake hands, and enter the interview room.
15. Go through the interview, asking and responding to questions.
16. Thank the interviewer, and leave the interview room.
17. Make notes about what happened during the interview.
18. Send a letter of thanks to the interviewer.
19. Wait to be notified of the firm's decision.
20. Follow up with a letter or phone call to the interviewer if not notified of the firm's decision within a reasonable interval.

APPENDIX B

Irrelevant Actions

Have lunch with the firm's representative.

Send college transcripts to the firm's Personnel Director.

Send an introductory letter to the Personnel Director requesting an interview.

Negotiate terms of employment.

Make travel arrangements to visit the firm.

Follow up the letter of introduction with a phone call to the Personnel Director to make an appointment for an interview.

Contact former professors and/or employers to send letters of recommendation to the Personnel Director.

Travel to the firm's headquarters for the interview.

Contact the firm to get the name of the Personnel Director or Manager of Executive Placement.

Check into a hotel the night before the interview.

Compose a letter of introduction to send to the firm.

Make an appointment with the firm's Personnel Director to visit the company.

Be met at the airport by the firm's representative.

Spend the day meeting with various executives of the firm and touring the facility.

Receive a request from the Personnel Office to send recommendations from former professors and/or employers.

REFERENCES

Abbott, Valerie and John B. Black (1980), "The Representation of Scripts in Memory," *Cognitive Science Technical Report #5*, New Haven, CT: Yale University Press.

Abelson, Robert P. (1976), "Script Processing in Attitude Formation," in *Cognition and Social Behavior*, eds. John S. Carroll and John W. Payne, Hillsdale, NJ: Lawrence Erlbaum, 33–45.

Alba, Joseph W., Susan G. Alexander, Lynn Hasher, and Karen Caniglia (1981), "The Role of Context in the Encoding of Information," *Journal of Experimental Psychology: Human Learning and Memory*, 7 (4), 283–292.

——— and Lynn Hasher (1983), "Is Memory Schematic?" *Psychological Bulletin*, 93 (March), 203–231.

Bartlett, Sir Frederick (1932), *Remembering*, Cambridge, England: Cambridge University Press.

Black, John B. and Hyman Bern (1981), "Causal Coherence and Memory for Events in Narratives," *Journal of Verbal Learning and Verbal Behavior*, 20 (3), 267–275.

Bower, Gordon H., John B. Black, and Terrence J. Turner (1979), "Scripts in Memory for Text," *Cognitive Psychology*, 11 (2), 177–220.

Bransford, John D. and Marcia K. Johnson (1972), "Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall," *Journal of Verbal Learning and Verbal Behavior*, 11 (December), 717–726.

Churchill, Gilbert A., Jr. (1979), "A Paradigm for the Development of Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64–73.

Ericsson, K. Anders and Herbert A. Simon (1980), "Verbal Reports as Data," *Psychological Review*, 87 (3), 215–251.

Ferber, Robert (1949), *Statistical Techniques in Market Research*, New York: McGraw-Hill.

Graesser, Arthur C., Sallie E. Gordon, and John D. Sawyer (1979), "Recognition Memory for Typical and Atypical Actions in Scripted Activities: Tests of a Script Pointer + Tag Hypothesis," *Journal of Verbal Learning and Verbal Behavior*, 18 (June), 319–332.

———, Stanley B. Woll, Daniel J. Kowalski, and Donald A. Smith (1980), "Memory for Typical and Atypical Actions in Scripted Activities," *Journal of Experimental Psychology: Human Learning and Memory*, 6 (5), 503–515.

Holsti, Ole R. (1969), *Content Analysis for the Social Sciences and Humanities*, Reading, MA: Addison-Wesley.

John, George and John C. Whitney (1982), "An Empirical Investigation of the Serial Structure of Scripts," *AMA Educators' Conference Proceedings*, Chicago: American Marketing Association, 75–79.

Kassarjian, Harold (1978), "Presidential Address 1977: Anthropomorphism and Parsimony," in *Advances in Consumer Research,* Vol. 5, ed. H. Keith Hunt, Ann Arbor, MI: Association for Consumer Research, xiii–xiv.

Lichtenstein, Edward H. and William F. Brewer (1980), "Memory for Goal-Directed Events," *Cognitive Psychology,* 12 (3), 412–445.

Mandler, George (1982), "The Structure of Value: Accounting for Taste," in *Affect and Cognition,* eds. Margaret S. Clark and Susan T. Fiske, Hillsdale, NJ: Lawrence Erlbaum, 3–36.

Martin, Joanne, Wendy Harrod, and Caren Siehl (1980), "The Development of Knowledge Structures," Research Paper No. 557, Stanford University.

Nelson, Katherine (1981), "Social Cognition in a Script Framework," in *Social Cognitive Development,* eds. Lee Ross and John H. Flavell, Cambridge, England: Cambridge University Press, 97–118.

Nisbett, Richard E. and Timothy D. Wilson (1977), "Telling More Than We Know: Verbal Reports on Mental Processes," *Psychological Review,* 84 (3), 231–259.

Olshavsky, Richard W. and Donald H. Granbois (1979), "Consumer Decision Making—Fact or Fiction?" *Journal of Consumer Research,* 6 (September), 93–100.

Rumelhart, David E. and Andrew Ortony (1977), "The Representation of Knowledge in Memory," in *Schooling and the Acquisition of Knowledge,* ed. Rand J. Spiro and William E. Montague, Hillsdale, NJ: Lawrence Erlbaum, 99–135.

Schank, Roger C. and Robert P. Abelson (1977), *Scripts, Plans, Goals and Understanding,* Hillsdale, NJ: Lawrence Erlbaum.

Whitney, John C. and George John (1983), "An Experimental Investigation of Intrusion Errors in Memory for Script Narratives," in *Advances in Consumer Research,* Vol. 10, eds. Alice M. Tybout and Richard P. Bagozzi, Ann Arbor, MI: Association for Consumer Research, 661–666.