Conceptions of the Psychological Research Process: Script Variation as a Function of Training and Experience

DOUGLAS A. HERSHEY
George Mason University

THOMAS L. WILSON
Bellarmine College

JENNIFER MITCHELL-COPELAND
George Mason University

Individuals' mental representations of the psychological research process were investigated. One-hundred eight subjects, representing four different levels of prior training in psychology (undergraduate students to full professors), individually generated a list of events involved in the psychological research process. Content analyses of the lists revealed high levels of consensus for specific events that are central to the research process (e.g., design experiment, collect data). Descriptive analyses of group differences identified developmental trends in both the number and types of events that were generated. The data support the notion that individuals in all four groups possessed *scripts* of the psychological research process. An expert script of the research process is also presented. Findings are discussed in terms of their implications toward a psychology of science.

Knowledge of a general set of procedures to follow when conducting scientific research is essential to the work of psychologists. In classrooms and laboratories countless hours are spent teaching students to understand the research process. Through formal education and practical research experience, psychologists-in-training acquire the necessary tools to carry out scientific studies, and develop the knowledge and skills required to critically evaluate empirical reports. Given these facts, it is surprising that relatively little attention has been focused on psychologists' knowledge of the psychological research process (Estes, 1993), and its development.

The notion that psychologists should bring their specialized knowledge of human behavior to bear on the activities of scientists is not new (Maslow, 1966; Singer, 1972; Stevens, 1936, 1939). A number of individuals (e.g., T. Kuhn, 1970a; Mahoney, 1976; Tweney, Doherty, & Mynatt, 1981) have made persuasive arguments suggesting that science, as an enterprise, would benefit from knowledge of how researchers conceptualize their work activities. Along these lines, Mahoney (1976) makes the point that psychologists have neglected to study the primary instrument of all research—the

Current Psychology: Developmental • Learning • Personality • Social Winter, 1996, Vol. 14, No. 4, 293–312.

researcher. Other authors have also pointed out the need for psychologists to examine scientists' conceptions of the research process (Gholson & Houts, 1989; Gorman & Carlson, 1989). This general form of research (i.e., science-on-science) is said to be *reflexive* in that the same techniques and methodologies used in the laboratory are systematically applied to study the activities of those who conduct the scientific work (Barker, 1989; Shadish, Fuller, & Gorman, 1994).

Although it was recently proclaimed that "the psychology of science has finally arrived" (Shadish et al., 1994, p. 3), the vast majority of work in this area has focused on the issue of "whether or not scientists behave according to the canons of rationality outlined by philosophers" (Shadish & Neimeyer, 1989). A small number of empirical studies have recently been conducted, however, which focus on how scientific thinking skills develop (Kuhn, Amsel, & O'Loughlin, 1988). The majority of this work has been based on one of two qualitatively different approaches to this issue. The first major approach has been to analyze the reasoning strategies that non-scientists use when confronted with a novel problem-solving task. This approach is exemplified by the work of Klahr and Dunbar (1988) who trained subjects to operate a fairly complex model tank, and then required them to *discover* how to make the tank perform functions that were not included in the training. In-depth analyses of subjects' thought processes provided these investigators with sufficiently rich data to formulate a comprehensive theoretical model of the scientific reasoning process.

The second general approach used to understand how scientific thinking skills develop has involved examining the thought processes of scientists responsible for significant scientific advancements. This approach is exemplified by the work of Kulkarni and Simon (1988), who provide a detailed analysis of the heuristics and reasoning processes that led biochemist Hans Krebs to discover the urea cycle in 1932. Similar retrospective analyses were used by Weisberg (1986) in his examination of the creative process that led mathematician Henri Poincare to discover a complex set of geometric functions, and Tweney's (1985, 1989) examination of the "scripts" Michael Faraday used to conduct his classic experiments on electromagnetism.

Evidence from these two separate lines of study (the "layperson" approach, and the "great scientific discovery" approach) suggests that the scientific research process can be divided into at least three qualitatively different sets of conceptually oriented activities: hypothesis generation, experimentation, and inference (cf., Kuhn et al., 1988). What these studies fail to reveal, however, is information regarding the concrete set of activities that researchers engage in while conducting experimental research. Herbert Simon points out that philosophers have made serious attempts to address the issue of how research is conducted; however, they often miss the mark by describing "how scientists *ought* to proceed, in order to conform with certain conceptions of logic, [rather] than with how they [actually] *do* proceed [in conducting scientific research]" (1977, p. 287, emphasis in original). In the present study an attempt was made to fill this apparent gap in the existing literature by examining the set of concrete activities in which psychologists are engaged during the course of a typical scientific investigation.

According to Gholson, Freedman, and Houts (1989; see also Gholson & Houts, 1989), "a major goal for the cognitive psychology of science is to provide a cognitive theory that can account for how the working practices of scientists lead[s] to . . . scientific knowledge." (p. 267). As an initial step toward that objective, the primary goal of the present study was to examine developmental differences in psychologists' knowledge of the research *process.* A cognitive developmental framework was adopted in order to address the following questions: What, if anything, does an undergraduate know about the psychological research process prior to completing his or her first course in the field? What do students learn about conducting psychological research during their undergraduate training? How does the knowledge of recent Ph.D.'s differ from that of graduate students? And finally, to what extent does knowledge of the research process differ between junior and senior professors who are actively engaged in conducting research?

The present study is unique in that the focus of the psychological research spotlight is directed upon psychologists, in order to gain a better understanding of developmental differences in knowledge of the research process. In doing so, the study revealed some interesting findings about how psychologists conceptualize their research endeavors. However, before outlining our methodological approach, we provide a brief overview of the theoretical assumptions that underlie the investigation.

Theoretical Framework

Psychologists' knowledge of the sequence of activities involved in conducting experimental research appears to be appropriately characterized as a script. Scripts, which are one type of procedural knowledge representation, have been described as a set of ordered actions that are linked together in long-term memory (Abelson, 1975; Schank, 1975; Schank & Abelson, 1977). One might think of a script as a "compiled" mental event sequence, containing those activities typically associated with a commonly experienced event. In one of the first empirical tests of the script hypothesis, Bower, Black, and Turner (1979) asked subjects to list the set of actions that are typically associated with going to a restaurant. These researchers found a striking degree of overlap in the specific actions subjects mentioned, which they took as evidence for the hypothesized event-based knowledge structure. Events in a script are assumed to be a good sample of the most available actions in long-term memory (Tversky & Kahneman, 1973), and the frequency of mention of an event across subjects is assumed to be an index of the salience or centrality of that item (Galambos, 1986). Since the introduction of the script construct, individuals have been shown to possess scripts for a variety of simple, commonly experienced events such as washing a car (Graesser, Woll, Kowalski, & Smith, 1980), doing the grocery shopping (Light & Anderson, 1983), and even using a toothbrush (Greene, Houston, Reinsmith, & Reed, 1992).

More recently, the script construct has been adopted to describe experts' knowledge of more complex event-based situations. For example, Hershey, Walsh, and their colleagues have suggested that practice at solving complex problems leads to the acquisition of problem-solving scripts (Hershey, Walsh, Read, & Chulef, 1990; Walsh & Hershey, 1993). They proposed that a problem-solving script not only specifies a set of information that should be considered in a particular problem-solving context, but the script also specifies an optimal sequence in which the information should be processed (Hershey, Morath, & Walsh, 1991; Kuboi, Hershey, & Walsh, 1990). These researchers further argued that the quality of one's domain-specific scripts within fields characterized by complex problem solving (e.g., legal adjudication, retirement planning, auto mechanics) is an important determinant of an individual's level of expertise. This is because the strategy-laden scripts of experts, which have been pruned and refined on the basis of the outcomes of numerous past performances, lead to the production of high-quality solutions based on a minimal amount of effort. Others have argued that the activities associated with conducting scientific research constitute a unique form of complex problem solving (T. Kuhn, 1970b; Simon, 1977). Taken together, these arguments suggest to us that an experienced researcher is likely to possess a detailed, event-based representation of the psychological research process.

It has been proposed that scripts, like other forms of knowledge structures, are hierarchically organized in long-term memory (Abbot, Black, & Smith, 1985; Galambos, 1986; Graesser, 1978; Lichtenstein & Brewer, 1980). Bower et al. (1979) concluded that a "script is not a linear chain of actions at one level, but rather a hierarchically organized 'tree' of events with several levels of subordinate actions" (p. 186). These authors went on to suggest that such hierarchies are really "goal trees" that contain multiple superordinate and subordinate sets of behavioral goals. Abelson (1981) coined the term "metascript" to describe the set of goals that are superordinate to the set of events contained in a simple script. He further suggested that these metascripts can serve as higher-order goal structures for *multiple* lower-level scripts. For example, a script of the psychological research process might be thought of as just one of many different scripts that are used to guide research based upon the principles of the scientific method. That is, an individual who has a psychological research script is likely to possess some form of basic-level script (good or poor) for how to conduct experiments in chemistry, biology, and physics. It is further conceivable that at the next higher level in the experimental goal tree there exists a "general scientific research metascript" that would contain goals common to research across all four of these disciplines. It may be that these higher-order goals include hypothesis generation, experimentation, and inference, the conceptually based activities found in the research on scientific reasoning previously discussed.

In the present study we examined how individuals' scripts of the psychological research process differed as a function of expertise. The scripts of four groups of subjects with different amounts of prior research training and practical research experience were compared. These scripts were obtained by asking subjects to make a list of the typical actions, steps, or stages that a psychologist would engage in when working on a research problem. We then used these event lists to compare the psychological research scripts of the four groups. Toward this end, we conducted both qualitative and quantitative analyses of the data, using a variety of dependent measures to characterize developmental trends. We not only sought to identify developmental differences in subjects' scripts, but we also had our subjects decompose their event lists into a subset of conceptually based, higher-order categories in order to identify a research metascript for the psychological research process. Finally, we present an empirically derived "expert" script of the research process based on the events listed by the forty-nine psychology professors in the sample.

METHOD

Participants

Four groups of subjects with different amounts of experience in psychological research were sampled: (a) undergraduate students enrolled in an introductory psychology course [n = 31], (b) first-year psychology doctoral students [n = 28], (c) assistant professors of psychology holding tenure-track appointments [n = 231, and (d) full professors of psychology [n = 26]. Members of the undergraduate group were students attending a large suburban university. Subjects in the other three groups were graduate students and faculty members sampled from among seven different major East Coast universities.

Graduate students and faculty represented a variety of areas within the field of psychology: Biological, Clinical, Counseling, Cognitive, Developmental, Human Factors, Industrial/ Organizational, and Social. All faculty members who participated in the study were actively involved in psychological research; assistant professors reported spending an average of 57% of their job-related effort on research activities, and full professors reported that 41% of their total effort was devoted to research. Undergraduates participated in the study for partial course credit, and all other subjects voluntarily participated in the study without remuneration.

Procedure

Subjects were asked to list about twenty actions, steps, or stages that characterize the process psychologists go through when working on a research problem. They were told that we were interested in their impression of the typical actions that a psychologist would engage in while carrying out a psychological research project. In order to establish common anchor events for subjects at both ends of the event sequence, the event "*Get idea for project*" was printed at the top of the response form, and "*Publish the research paper*" was printed at the bottom of the sheet. Upon completion of the event list, subjects were asked to review their work to ensure that important steps had not inadvertently been omitted, and all events listed were ordered in the correct sequence. Following this review subjects were allowed to modify their list as necessary. Subjects were then asked to draw lines between events in order to divide the script into a smaller number of higher-order categories, and finally they were asked to write down names for each of the categories they created. This method for identifying the higher-order goal structure of the script is conceptually similar to the technique described in Galambos (1986; Experiment 6).

Scoring the Research Scripts

The first step in the scoring process involved compiling a comprehensive master list of unique research activities from among the 1507 events subjects generated.[1] The resulting 103 events are presented as an appendix, listed in the serial order in which they were commonly mentioned. A subset of these events appeared to be subordinate to other events. For example, subjects listed graph data, descriptive statistics, inferential statistics, and data analysis. Rather than scoring each of these unique items as instances of data analysis, we chose to include all four events on the master list. This was done in an effort to maintain the richness inherent in the data. All "subordinate items," such as the three former items listed are indented on the list of activities contained in the Appendix.

After the master list of events had been compiled, it was used to score individuals' research scripts. Trained raters matched events listed on each subject's script with activities appearing on the master list. Prior to scoring, the protocols were transcribed and group-linked identifiers were removed from the scripts so as not to influence raters' knowledge of a subject's group status. In order to establish interrater reliability for the scoring method, eight of the 108 protocols (two randomly selected from each of the four groups) were each scored by three different raters. Twenty-four calculations were made of the percentage of items identically chosen across pairwise combinations of raters. The median level of interrater agreement was found to be 92% across all pairwise combinations of raters.

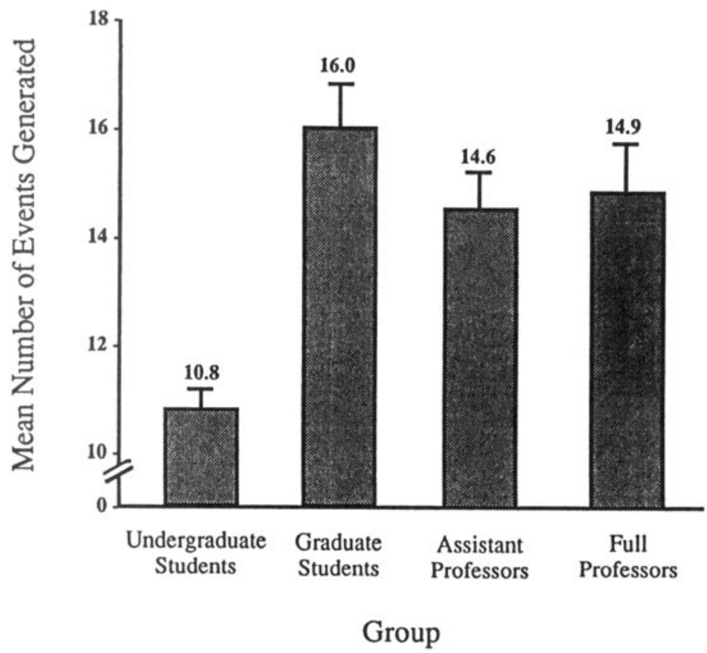
RESULTS

We began our analysis of the research scripts with a simple count of the number of events subjects generated. We asked subjects to list about twenty actions or events involved in the research process; however, we found substantial variability in the number of events subjects actually mentioned. A large majority of individuals generated fewer than the requested twenty items. Across all 108 subjects, the median number of items generated was 15, the minimum was 5, and the maximum was 23. Although in actuality there are far more than twenty discrete events that take place during the course of a psychological study, the above data suggest that a set of approximately 15 items are sufficient to characterize the subset of major events involved in a typical research project.

Based on the above finding, we sought to determine whether there were significant differences in the mean number of events generated by members of the four groups (see Figure 1).[2] A univariate analysis of variance revealed a significant main effect for the number of items generated across groups, $F\ [3, 104] = 12.10, p < .01, MS_e = 12.37$. An inspection of the means in Figure 1 shows that undergraduates generated the fewest events, graduate students listed the greatest number of events, and the two professor groups listed equivalent numbers of items. Pairwise planned comparisons indicated that undergraduates generated significantly fewer events than each of the other three groups: graduate students, $t\ [57] = 5.79, p < .01$; assistant professors, $t\ [51] = 4.95, p < .01$; and full professors, $t\ [56] = 4.71, p < .01$. No other significant differences were found in the remaining pairwise comparisons of list length.

We also compared the number of items each group collectively reported from among the universe of 103 items contained in the master list. Our working assumption here was that the number of items reported among members of a group is indicative of the

FIGURE 1
Mean number of events generated by each of the four groups (with standard error bars shown).



Undergraduates generated significantly fewer events than did subjects in the other three groups ($p < .05$); all other pairwise comparisons failed to obtain significance.

group's collective breadth of knowledge about the full set of relevant research activities. Among undergraduates, 58 of the total pool of 103 events were identified. In contrast, graduate students collectively identified 80 of the 103 events, and both of the professor groups identified 76 of the 103 events. To determine whether there was a statistically significant difference between the undergraduates and the other three groups, a z-test for independent proportions was performed between undergraduates and the group with the next larger proportion, assistant professors (i.e., comparing 58 to 76 items). The test revealed that undergraduates generated significantly fewer events than assistant professors, $z = 2.66, p < .01$.

Given the above developmental differences in the number of script events reported, we turned our attention to the particular events in each group's scripted knowledge. Four composite research scripts were formed, each on the basis of those events identified by at least twenty percent of subjects within a group. This minimum criterion for including an event in a composite script allowed us to eliminate relatively idiosyncratic items reported by only a few members of a group, and at the same time create group-based representations that would not differ appreciably from individual subjects' representations at each of the four developmental levels. It was our working assumption that those items listed most frequently by a group represent the most salient events in that groups' conception of the research process. We wish to emphasize that we do not assume that two individuals who mention the same event have comparable knowledge of that event. For example, an undergraduate who mentions data analysis is not likely to have an understanding of that activity equal to that of a full professor.[3]

The four composite scripts shown in Table 1 are presented using the same basic format employed by Bower et al. (1979). High-consensus (HC) events, which were named by ≥ 60% of the members of a group, are displayed in capital letters. Moderate consensus (MC) events, identified by 40–59% of members within a group, are displayed in upper and lowercase letters. And low consensus (LC) events, selected by 20–39% of subjects, are displayed in italics. These composite scripts formed the basis for examining the reliability of the group frequency norms for particular events and group differences in consensus.

To determine whether members of a group agree on the events they reported, each group was divided in half, and then the frequencies of mention of specific events by the two halves were correlated.[4] The Pearson correlation coefficients for the four groups were: undergraduates, $r = .38$, $p < .05$; graduate students, $r = .71$, $p < .01$; assistant professors, $r = .77$, $p < .01$; and full professors, $r = .84$, $p < .01$. The statistical significance of each of these correlations suggests a high reliability in the frequency with which groups reported particular script events. Furthermore, there is a monotonic increase in the reliability of the group frequency norms as a function of experience, with the greatest increase occurring between the undergraduate and graduate students.

With respect to group differences in consensus, Table 1 also reveals a dramatic increase in within-group agreement as one moves from the scripts of undergraduates to those of graduate students. The undergraduate composite script contains mainly LC events, a few MC events, and only two HC events. In contrast, the graduate student composite script contains seven HC events, two more MC events than undergraduates, and fewer LC events. Interestingly, no comparable increase in frequency of mention was found when comparing the consensus rates of the three more advanced groups. These differences in consensus suggest that the most significant development of the script occurs as individuals receive their undergraduate training. By the time students enter graduate school, they are in high agreement as to which specific events are important components of the research process.

In an effort to quantify the degree of between-group similarity in the events generated, the percentages of mention of all 103 events by all pairwise combinations of groups were correlated. These Pearson correlations represent the degree of relationship between the frequency of mention of events found for one group and the frequency of mention of events for another. The six pairwise between-group correlations are shown in Table 2. Each of these coefficients are significantly different from zero, suggesting that in spite of any apparent group-based differences in particular elements of the scripts, there is significant correspondence among the groups in terms of the types of events that were mentioned. The largest correlations were found among the three more advanced groups, and slightly smaller correlations were found between undergraduates and the other groups. These findings are important because they provide evidence that the scripts for the four groups are more similar than they are different. However, the

TABLE 1

Composite scripts for each of the four groups after eliminating activities that generated within-group agreement levels of less than 20%.

Undergraduate Students	Graduate Students	Assistant Professors	Full Professors
Get Idea (anchor)	Get Idea (anchor)	Get Idea (anchor)	Get Idea (anchor)
Locate Literature	*Locate Literature*	READ LITERATURE	*Locate Literature*
Go to Library †	READ LITERATURE	Discuss Idea	READ LITERATURE
READ LITERATURE	*Critically Evaluate Idea* †	*Conceptualize Project*	*Discuss Idea*
Organize Notes from Lit. †	*Discuss Idea*	Formulate Hypotheses	*Conceptualize Project*
Discuss Idea	Conceptualize Project	DESIGN EXPER. METHODS	*Determine Subject Population* †
Formulate Hypotheses	FORMULATE HYPOTHESES	*Obtain Materials/Measures*	Formulate Hypotheses
Design Exper. Methods	DESIGN EXPER. METHODS	*Construct Materials/Measures*	DESIGN EXPER. METHODS
Obtain Materials/Measures	*Construct Materials/Measures*	PILOT TEST PROCEDURES	*Obtain Materials/Measures*
Obtain Subjects	*Write Proposal for Funding* †	Refine Experiment	*Obtain Research Assistants* †
Data Collection	*Submit Ethics Review Forms*	*Obtain Subjects*	*Submit Ethics Review Forms*
Code/Organize Data	*Prepare Exper. Environment*	DATA COLLECTION	*Prepare Exper. Environment*
Data Analysis	Pilot Test Procedures	*Code/Organize Data*	Pilot Test Procedures
Hypotheses Supported?	*Recruit Ss for Pilot Study* †	DATA ANALYSIS	*Refine Experiment*
Replicate Findings †	*Refine Experiment*	Hypotheses Supported?	Obtain Subjects
Outline Paper †	OBTAIN SUBJECTS	*Make Presentation*	DATA COLLECTION
WRITE DRAFT	DATA COLLECTION	*Final Literature Review* †	Code/Organize Data
Check Grammar/Style †	*Data Entry* †	WRITE DRAFT	DATA ANALYSIS
Get Feedback	DATA ANALYSIS	*Get Feedback*	*Make Presentation*
Revise Draft	Hypotheses Supported?	*Revise Draft*	WRITE DRAFT
Publish Paper (anchor)	WRITE DRAFT	Submit for Publication	*Get Feedback*
	Get Feedback	*Make Post-review Revisions*	*Revise Draft*
	Revise Draft	Publish Paper (anchor)	Submit for Publication
	Submit for Publication		Publish Paper (anchor)
	Make Post-review Revisions		
	Publish Paper (anchor)		

High-consensus events (identified by ≥ 60% of subjects in a group) are shown in capital letters. Moderate-consensus events (identified by 40–59% of subjects in a group) are shown in upper and lowercase letters. Low-census events (identified by 20–39% of subjects in a group) are shown in italics. Events unique to a particular group are indicated by a dagger (†).

fact that the correlations are less than unity suggests that a small subset of items in the script change as individuals gain task-related experience.

Another question is whether the temporal sequence in which subjects listed events also show a high degree of similarity across groups. One of the defining characteristics of scripts is that they are sequentially organized. The sequential integrity of subjects' scripts was examined in an effort to determine whether scripts become increasingly well organized as a function of research experience. Upon completion of the task, subjects were given an opportunity to review their event lists in order to insert events that had been omitted, and rearrange events into what they believed to be the correct sequence. All instances of these omitted and misordered events were identified and coded as sequence errors.[5] Fully one-third of the sample (32.4%) was found to have made one or more sequence errors. Furthermore, the number of subjects who made sequence errors differed as a function of expertise. Forty-two percent of undergraduate students and 39% of graduate students made sequence errors. In contrast, only 22% of assistant professors and 23% of full professors made sequence errors. Related analyses, which focused solely on the 35 subjects who made sequence errors, revealed an inverse relationship between expertise and the number of errors generated. Undergraduates ($n = 13$) averaged 3.8 sequence errors, graduate students ($n = 11$) made 3.0 errors, assistant professors ($n = 5$) averaged 1.8 errors, and full professors ($n = 6$) made 1.3 errors.

In an effort to identify the metascript for the psychological research process, we asked our subjects to divide their event lists into a smaller set of higher-order categories, and provide names for each of the categories they created. We found that subjects readily carried out this request. There was a surprisingly high level of consistency in the number of categories subjects identified, suggesting a common core of higher-order goals in the research metascript. Eighty percent of subjects divided their lists into between three and five categories. The modal number of categories created was four, which was generated by fully one-third of the sample. The mean number of categories created for each of the four groups is depicted in Figure 2. The relatively small standard errors associated with these means suggest that there was high consensus regarding the number of higher-order categories in the metascript. A univariate analysis of variance revealed a significant main effect for the number of categories generated across groups, F [3, 104] = 3.99, $p < .01$, $MS_e = 1.38$. Pairwise planned comparisons indicated that graduate students generated significantly more categories than each of the other three groups: undergraduates, t [57] = 2.68; $p < .01$; assistant professors, t [48] = 2.32, $p < .05$; and full professors, t [53] = 2.95, $p <.01$. No other significant differences were found in the remaining pairwise comparisons of the number of categories generated.

Examination of the category labels revealed that the four most frequently cited labels outline a general sequence that defines the scientific method: (a) formulate ideas, (b) collect data, (c) analyze data, and (d) report findings. Although the precise terms subjects used to specify these higher-order goals varied from one individual to the next, the semantic content of the labels was always clear, and in each case where all four categories were cited, they were ordered in the cannonical sequence previously

TABLE 2
Between-group correlations for the 103 events contained on the master list.

	Group			
Group	Undergraduate Students	Graduate Students	Assistant Professors	Full Professors
Undergraduate Students	1.00			
Graduate Students	.71	1.00		
Assistant Professors	.69	.88	1.00	
Full Professors	.69	.91	.91	1.00

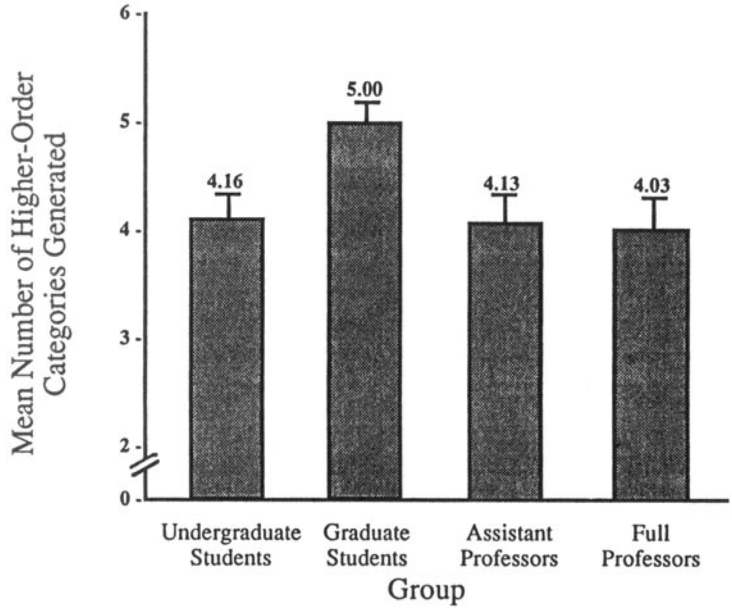
All correlations are significantly different from zero at the .01 level.

outlined. Moreover, this consistency in labeling and sequencing did not vary across groups, providing further evidence that these four higher-order goals in the research process constitute a general experimental metascript.

Finally, we compiled an expert script based on the set of most frequently cited events mentioned by the 49 subjects in the sample who held the Ph.D. (i.e., assistant and full professors). Table 3 contains an empirically derived expert psychological research script based solely on the set of events mentioned by these subjects. The minimum twenty percent agreement criteria (previously described in relation to Table 1) was again used to determine whether an event would be included in the expert script.

It is interesting to note that the five high-consensus events in this 23-item expert script (read literature, design experimental methods, data collection, data analysis, and write draft of paper) correspond rather well to the set of four category labels identified in the metascript analysis recently reported (formulate idea, data collection, data analysis, and report findings). The first two HC events in the expert script, read literature and design experimental methods are both important activities that take place during the early stages of the research process, as the researcher is formulating and refining his or her ideas. The next two HC events in the expert script (data collection and data analysis) actually match the labels most subjects used to describe two of the higher-order goals in the research process. Finally, the last HC event, write draft of paper, is perhaps the single most important behavior the researcher engages in in order to report the findings of the experiment to the scientific community.

FIGURE 2
Mean number of higher-level categories generated by each of the four groups (with standard error bars shown).



Graduate students generated significantly more categories than did subjects in the other three groups ($p < .05$); all other pairwise comparisons failed to obtain significance.

DISCUSSION

Scripts of the Psychological Research Process

Three pieces of evidence suggested that individuals in all four groups possessed scripts of the research process. First, subjects were able to generate event lists quickly and easily when requested to do so, without requiring additional instructions or clarifications. The ease with which subjects carried out the task suggests that they were reporting events from a preexisting sequentially organized representation. Second, across all groups, subjects showed high levels of agreement on a subset of basic level events that are particularly important to the research process (e.g., read literature, design experimental methods, data collection). Third, we found that subjects were readily able to decompose their event lists into a reduced set of higher-order goals, providing evidence that the scripted actions were subordinate to a superordinate goal structure (Abelson, 1981; Bower et al., 1979; Lichtenstein & Brewer, 1980; Mandler, 1979). These three characteristics of the psychological research script are consistent with the characteristics of other scripts that have been reported in the literature (Brewer, 1987; Mandler, 1979).[6]

Analyses of the higher-order goals generated by subjects provided additional insight into how individuals conceptualize the psychological research process. For example,

TABLE 3
A composite psychological research script based on those events that were mentioned by a minimum of 20% of the 49 professors in the sample.

Get Idea (anchor)
READ LITERATURE
Discuss Idea
Conceptualize Project
Determine Subject Population
Formulate Hypotheses
DESIGN EXPERIMENTAL METHODS
Obtain Materials/Measures
Construct Materials/Measures
Obtain Research Assistants
Pilot Test Procedures
Refine Experiment
Obtain Subjects
DATA COLLECTION
Code/Organize Data
DATA ANALYSIS
Hypotheses Supported?
Make Presentation
Final Literature Review
WRITE DRAFT
Get Feedback
Revise Draft
Submit for Publication
Make Post-review Revisions
Publish Paper (anchor)

High consensus events (shown in capitals) were mentioned by ≥ 60% of professors, moderate consensus events (shown in upper and lowercase letters) were mentioned by 40–59% of professors, and low consensus events (shown in italics) were mentioned by 20–39% of professors.

there was relatively little variability found in the number of goals generated, suggesting that the superordinate goal structure for publishing an empirical report is based on a four-step process (plus or minus one). This structure includes formulating ideas, collecting data, analyzing data, and reporting findings, which correspond to the set of goals other researchers have identified in studies of the scientific reasoning process. In summarizing that literature, Kuhn et al. (1988) concluded that the scientific research process involves three stages: hypothesis generation, experimentation, and inference. These three stages correspond to three of the superordinate goals identified by our subjects: formulate ideas, collect data, and analyze data, respectively. The fourth higher-order goal in the psychological research script, report findings, appears to be an artifact of the specific anchor event we employed (i.e., publishing the research paper).

Developmental Differences in the Scripts

A primary goal of the present study was to examine how the content of scripts of the psychological research process differs as a function of formal academic training and experience in conducting research. A cursory inspection of the undergraduates' composite script revealed a fairly representative sample of the set of activities involved in the psychological research process. On average, they were able to identify eleven steps in the process, and there was at least moderate levels of agreement for a significant number of actions. A closer inspection of their script, however, indicated that they lacked a depth of understanding regarding the finer points of experimentation. As a group, undergraduates only mentioned about half of the total pool of elements contained in the master list, which, in comparison to the other three groups, was a significantly smaller subset of events. The split-group reliability analysis of the undergraduates' scripts revealed that their event lists were reliable; however, they were substantially less reliable than the scripts generated by members of the other three groups. Furthermore, undergraduates mentioned fewer moderate and high-consensus events than the more advanced groups.

In contrast, the metascripts of the undergraduates were found to be equivalent to those of the more advanced groups. Specifically, the number of higher-order goals undergraduates generated, and the labels they provided for those goals, were in high agreement with the number and labels of goals provided by the graduate students and professors. This finding suggests that it was the undergraduates' general scientific research metascript, presumably acquired in high school science classes, that allowed them to flesh out a reasonable script of the psychological research process. We speculate that this metascript facilitates the learning process by providing a rudimentary cognitive framework (Brewer, 1987) that guides the organization of newly acquired lower-level events.

Analysis of the graduate students' scripts suggests that psychology majors acquire a great deal of information about the research process during the course of their undergraduate training. This conclusion is based in part on the change in the within-group consensus rates for events found in the undergraduate and graduate student composites. Specifically, a number of the LC events in the undergraduate script were MC or HC events in the graduate student script. In addition, relative to all other groups, the graduate students' composite script was the longest, and they identified significantly more higher-order goals. These findings reflect the fact that graduate students' scripts were more richly defined than those of the other three groups. Taken together, these results suggest that the psychological research script is firmly established by the time students are admitted to a doctoral program.

In many respects the graduate students' scripts were indistinguishable from those generated by professors. The number of items they identified was not significantly different from the professors, the reliability analyses were comparable for the three advanced groups, and the between-group correlations of events revealed that activities graduate students listed were quite similar to events mentioned by assistant and full professors. One important difference between the scripts of the students and professors, however, became apparent in the analysis of sequence errors. Students' initial lists of the research process were found to be more poorly organized, and had more elements missing than the event lists generated by professors. Anderson's (1987) model of skill acquisition provides a plausible explanation for this finding. In that model, it is suggested that through task-specific experience, individual procedures (or events, in the present context) become "composed" into larger chunks of procedures as part of the knowledge compilation process. The more experienced subjects were found to have made fewer sequence errors, ostensibly due to the fact that their mental representations of research-related events had been composed into multievent procedural sequences.

The two professor groups were found to generate significantly fewer higher-order goals than graduate students, and significantly more events than undergraduates. However, the scripts of the assistant and full professors are best characterized by stability, rather than change. The composite scripts for the two professor groups were quite similar in terms of length, content and consensus (see Table 1). These findings suggest that research experience beyond graduate school does little to modify the basic representation of the script.

The Expert Script

The expert composite script presented in Table 3 is not offered as a prescriptive model of the research process. Rather, we present this script because we believe that it can serve as a useful pedagogical tool for training students to understand the research process. An inspection of the books on our shelves revealed that most Introductory Psychology and Research Methods texts present prototypical models of the research process. Many of the models, particularly those found in introductory texts, cover fundamentals of the scientific method (i.e., make hypotheses, make observations, form theories). Texts that were found to present reasonably comprehensive, procedural representations of the psychological research process were the exception, rather than the rule (although, see Runkel & McGrath, 1972 for a notable exception). None, however, were found to contain a model that approached the scope and complexity of the expert script. In our own introductory psychology and research methods courses we have found that the first 19 steps of the expert script can be used to provide a conceptual overview of the research process. We have also found that classroom discussion and general interest in the topic is enhanced by having students generate their own psychological research scripts prior to the presentation of the expert script.

CONCLUSION

The script construct appears to provide significant heuristic value toward establishing how scientists conceptualize their research endeavors. However, the preliminary nature of the present study suggests that a great deal of additional work will be required before scientists' knowledge of the research process is well understood. A broader and more ambitious goal in this line of inquiry would involve establishing an empirically based, comprehensive theory of the psychology of science. We propose that in order to achieve that goal, at least two separate but related lines of future research are indicated; one that focuses on the psychology of psychological research, and a second line of work that takes a more interdisciplinary perspective by focusing on the psychology of science. The rationale for the former is that psychologists have a vested interest in better understanding the methods they use to conduct their science. The latter line of work could serve to test the generality of the former, and provide a broader understanding of the variety of empirical methods used across disciplines.

One limitation of the present study is that it is unclear just how much knowledge an individual actually possesses about an event, based solely on his or her ability to list that event. Future studies could focus on the "microscripts" associated with specific events, and the psychological mechanisms that underlie the development of those specific knowledge structures. For example, one could conduct a detailed examination of students' and experts' knowledge of sampling procedures or data analysis. Further research at a finer level of analysis could provide insights into individual differences in students' understanding of research activities. It would also be informative to know whether students who are taught the psychological research script as a single coherent structure excel at scientific reasoning relative to peers who have acquired comparable knowledge in a traditional, piecemeal fashion. If this is found to be the case, then it suggests that we should rethink the way that we currently train students to become scientists. Studies such as these could be quite valuable in designing educational programs for student researchers.

It was pointed out in the introduction that there exists a paucity of empirical work on the psychology of science. Of interest would be answers to such questions as: Where do good scientific ideas come from? What specific thinking skills facilitate good science, and in turn, lead to significant contributions to the literature? How do psychologists' conceptions of the research process compare to those of scientists in other disciplines? Questions such as these have been posed in the past, but answers have always been discussed in a philosophical context (cf., Platt, 1964). This is where cognitive experimental psychology stands to make a considerable contribution (Gholson & Houts, 1989; Gorman & Carlson, 1989). In sum, we propose that an empirically oriented psychology of science is an important and viable area of research that is only now in its infancy. Nearly two decades ago Mahoney suggested that "any appreciation or improvement in knowledge requires an understanding of the knower" (1976, p. 27). Now, nearly twenty years later, we urge researchers to heed Mahoney's call by establishing an empirically based cognitive psychology of science.

NOTES

The authors would like to acknowledge Ray Morath for his assistance in the collection and scoring of the data, and James T. Austin and Stephen Zaccaro for critical comments on the manuscript.

Portions of this paper were presented in June of 1993 at the annual meeting of the American Psychological Society in Chicago.

Accepted for publication: 11 July 1995. *Address correspondence to* Douglas A. Hershey, Department of Psychology, George Mason University, Fairfax, VA 22030–4444.

1. A small number of activities (fewer than 2%) were eliminated from the pool of items for one of two reasons—either the precise nature of the activity was ambiguous, or the activity was trivial or not *directly* related to the completion of the research project (e.g., turn on computer).
2. The group data are fully reproduced in the Appendix.
3. The investigation of individual differences in knowledge of lower-level activities associated with particular events goes beyond the scope of the present study.

4. For groups with an odd number of subjects, one subject was randomly chosen and eliminated from the analysis before the remaining subjects were randomly assigned to the two halves.
5. It is important to note that the term *error* in this context is not used to suggest that subjects did not know where the event should be ordered in the list. The fact that many subjects respecified the order of events bears this point out. Instead, the term is used to imply that subjects were unable to spontaneously recall the canonically correct sequence of events. As such, sequence errors reflect a difficulty in retrieving information from long-term memory.
6. In addition, like other scripts that have been described in the literature, scripted representations of the research process are not assumed to be algorithmic. Research is not a rigidly structured linear arrangement of activities. Of course, data must be analyzed before conclusions can be drawn, but many of the steps in the research process can be recursive.

REFERENCES

Abbott, V., Black, J.B., & Smith, E.E. (1985). Representation of scripts in memory. *Journal of Memory and Language, 24,* 179–199.
Abelson, R.P. (1975). Concepts for representing mundane reality in plans. In D.G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, pp. 273–309.
Abelson, R.P. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715–729.
Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review,* 94, 192–210.
Barker, P. (1989). The reflexivity problem in the psychology of science. In B. Gholson, W. R. Shadish, R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience.* New York: Cambridge University Press, pp. 92–114.
Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177–220.
Brewer, W. F. (1987). Schemas versus mental models in human memory. In P. Morris (Ed.), *Modelling cognition.* Chichester: Wiley, pp. 187–197.
Estes, W. K. (1993). Concepts, categories, and psychological science. *Psychological Science, 4,* 143–153.
Galambos, J. A. (1986). Knowledge structures for common activities. In J. A. Galambos, R. P. Abelson, & J. B. Black (Eds.), *Knowledge structures.* Hillsdale, NJ: Erlbaum, pp. 21–47.
Gholson, B., Freedman, E. G., & Houts, A. C. (1989). Cognitive psychology of science. In B. Gholson, W. R. Shadish, R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience.* New York: Cambridge University Press, pp. 267–274.
Gholson, B., & Houts, A. (1989). Toward a cognitive psychology of science. *Social Epistemology, 3,* 107–127.
Gorman, M., & Carlson, B. (1989). Can experiments be used to study science? *Social Epistemology, 3,* 89–106.
Graesser, A. C. (1978). How to catch a fish: The memory and representation of common procedures. *Discourse Processes, 1,* 72–89.
Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted events. *Journal of Experimental Psychology, 6,* 503–515.
Greene, T. R., Houston, S. E., Reinsmith, C. L., & Reed, E. (1992, November). *Evidence for scripts for everyday motor activities.* Paper presented at the annual meeting of the Psychonomic Society, St. Louis, MO.
Hershey, D. A., Morath, R., & Walsh, D. A. (1991, November). The use of scripts to solve financial planning problems. In C. Berg (Chair), *Age differences in cognitive strategies.* Symposium conducted at the meeting of the Gerontological Society of America, San Francisco.
Hershey, D. A., Walsh, D. A., Read, S. J., & Chulef, A. S. (1990). The effects of expertise on financial problem solving: Evidence for goal-directed problem solving scripts. *Organizational Behavior and Human Decision Performance, 46,* 77–101.
Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1–48.
Kuboi, A. S., Hershey, D. A., & Walsh, D. A. (1990, April). *Patterns of information search in problem solving.* Paper presented at the Western Psychological Association Conference, Santa Clara, CA.
Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* New York: Academic Press.
Kuhn, T. (1970a). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
Kuhn, T. (1970b). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge.* New York: Cambridge University Press, pp. 1–23.

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science, 12,* 139–175.
Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology, 12,* 412–445.
Light, L. L., & Anderson, P. A. (1983). Memory for scripts in young and older adults. *Memory and Cognition, 11,* 435–444.
Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative.* Cambridge, MA: Ballinger.
Mandler, J. (1979). Categorical and schematic organization in memory. In C. R. Puff (Ed.), *Memory organization and structure.* New York: Academic Press, pp. 259–299.
Maslow, A. H. (1966). *The psychology of science: A reconnaissance.* Chicago: Henry Regnery Co.
Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.
Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to methods.* New York: Holt, Rinehart & Winston.
Schank, R. C. (1975). The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding. Studies in cognitive science.* New York: Academic Press, pp. 273–309.
Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.
Shadish, W. R., Fuller, S., & Gorman, M. E. (1994). Overview. In W. R. Shadish & S. Fuller (Eds.), *The social psychology of science.* New York: Guilford Press, pp. 3–123.
Shadish, W. R., & Neimeyer, R. A. (1989). Contributions of psychology to an integrative science studies: The shape of things to come. In S. Fuller, M. De Mey, T. Shinn, & S. Woolgar (Eds.), *The cognitive turn: Sociological and psychological perspectives on science.* Boston: Kluwer, pp. 13–38.
Simon, H. A. (1977). *Models of discovery.* Boston: Reidel Publishing Company.
Singer, B. F. (1972). Toward a psychology of science. *American Psychologist, 26,* 1010–1016.
Stevens, S. S. (1936). Psychology: The propaedeutic science. *Philosophy of Science, 3,* 90–104.
Stevens, S. S. (1939). Psychology and the science of science. *Psychological Bulletin, 36,* 221–263.
Tweney, R. D. (1985). Faraday's discovery of induction: A cognitive approach. In D. Gooding & F. James (Eds.), *Faraday rediscovered. Essays on the life and work of Michael Faraday, 1791–1867.* New York: Stockton Press, pp. 189–210.
Tweney, R. D. (1989). A framework for the cognitive psychology of science. In B. Gholson, W. R. Shadish, R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience.* New York: Cambridge University Press, pp. 342–366.
Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (1981). General Introduction. In R. Tweney, M. Doherty, & C. Mynatt (Eds.), *On scientific thinking.* New York: Columbia University Press, pp. 1–23.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.
Walsh, D. A., & Hershey, D. A. (1993). Mental models and the maintenance of complex problem solving skills into old age. In J. Cerella & W. Hoyer (Eds.), *Adult information processing: Limits on loss.* New York: Academic Press, pp. 553–584.
Weisberg, R. (1986). *Creativity: Genius and other myths.* New York: W. H. Freeman & Company.

APPENDIX

Master List of Events

The following is a master list of 103 research activities representing all unique events generated by subjects. This master list was used by three blind raters to code each subject's list of events. Values in parentheses are the percentages of undergraduates, graduate students, assistant professors, and full professors listing each item. Indented events are items that were judged to be subordinate to the preceding nonindented event.

Get idea for project (anchor)
Observe phenomenon in real world (3; 0; 9; 0)
Obtain okay from advisor (3; 4; 0; 0)
Locate/obtain literature (39; 29; 9; 23)
    Go to library (23; 4; 4; 0)
Read literature on topic (74; 75; 91; 81)
Organize notes on idea (19; 14; 17; 0)
Cross-reference literature materials (0; 7; 0; 8)
Organize notes from literature (35; 11; 4; 8)
Observe presentations on topic (0; 0; 4; 0)
Critically evaluate research ideas (13; 32; 13; 15)
Discuss idea with other people (23; 25; 52; 27)
Get experienced collaborators for project (3; 7; 0; 4)
Formulate different theoretical conceptions (0; 4; 0; 8)
Conceptualize project (19; 46; 26; 35)
    Consider possible research methods (3; 14; 9; 15)
    Identify/conceptualize IV & DV (0; 18; 17; 8)
    Operationalize IV & DV (0; 14; 13; 8)
    Determine appropriate subject population (16; 14; 17; 23)
    Determine where to procure subjects (0; 4; 4; 8)
    Prepare outline of project (10; 4; 4; 4)
Formulate hypotheses (52; 68; 43; 46)
Link hypotheses to literature (0; 4; 4; 0)
Determine contributions to current literature (3; 4; 0; 8)
Refine hypotheses (0; 4; 9; 8)
Determine practicality of doing study (0; 4; 13; 12)
Determine possible confounds/bias (0; 11; 4; 8)
Design experimental methods (29; 71; 65; 69)
Obtain available experimental materials/measures (29; 11; 39; 23)
Construct materials/measures (10; 25; 26; 15)
Consider reliability of measures (0; 0; 0; 4)
Consider validity of measures (6; 4; 0; 0)
Determine sample size (power analysis) (0; 14; 0; 8)
Write proposal for funding (0; 21; 4; 12)
Receive funding for project (3; 18; 0; 4)
Search for additional sources of funding (0; 0; 4; 4)
Obtain research assistants (3; 11; 13; 27)
Train research assistants (0; 11; 13; 15)
Consider ethical issues (3; 4; 4; 4)
Develop debriefing information for subjects (0; 4; 4; 0)
Submit ethics review forms (0; 21; 9; 27)
Arrange computer needs (0; 4; 0; 0)
Prepare experimental environment (13; 21; 9; 27)
Pilot test procedures/measures (3; 43; 61; 42)
    Pilot test computer program (0; 0; 4; 0)
    Pilot test self (3; 0; 0; 12)
    Recruit subjects for pilot study (0; 21; 13; 12)

Analyze pilot data (0; 14; 9; 12)
Evaluate outcome of pilot study (0; 7; 17; 0)
Review pilot results with colleagues (6; 0; 0; 4)
Refine experiment based on pilot results (0; 29; 48; 27)
Advertise for subjects (6; 0; 4; 0)
Obtain subjects (39; 61; 30; 50)
Set testing schedule (3; 0; 4; 15)
Random assignment of subjects (0; 7; 0; 0)
Data collection (58; 100; 100; 96)
Observe subjects during experiment (13; 0; 0; 0)
Obtain subjects' perceptions of study (10; 0; 0; 0)
Debrief subjects (0; 14; 4; 8)
Backup computer data files (0; 0; 0; 4)
Code/Organize data (26; 14; 26; 42)
Data entry (10; 25; 17; 12)
Check data for errors & faulty recording (10; 11; 9; 19)
Data Analysis (23; 86; 91; 100)
Examine data (16; 4; 9; 12)
Graph data (0; 4; 4; 4)
Primary data analysis (descriptive statistics) (3; 11; 4; 8)
Advanced data analyses (inferential statistics) (10; 18; 17; 4)
Analyze reliability of measurement instrument (0; 0; 9; 0)
Determine if hypotheses were supported (45; 46; 48; 19)
Check for experimenter bias (3; 0; 0; 0)
Check for statistical artifacts (0; 4; 0; 0)
Consider validity of results (3; 4; 4; 0)
Perform unplanned data analyses (0; 11; 13; 4)
Review findings with statistician (0; 4; 0; 0)
Consider theoretical issues in light of data (0; 0; 4; 0)
Replicate experimental findings (29; 7; 9; 4)
Determine implications of research (0; 11; 4; 12)
Consider future experiments (0; 4; 9; 4)
Prepare a synthesis of findings (19; 0; 0; 12)
Make conference/brown bag presentation (16; 11; 22; 23)
Select journal for submission (0; 11; 4; 8)
Consider journal guidelines for proper style (0; 4; 4; 0)
Conduct final literature review (6; 4; 26; 15)
Outline paper (26; 0; 0; 4)
Write draft of paper (74; 82; 87; 81)
Write methods (0; 7; 4; 8)
Write results (3; 11; 4; 15)
Write introduction (6; 11; 4; 19)
Write discussion (6; 11; 4; 19)
Write abstract (0; 4; 0; 4)
Write references (16; 0; 0; 0)
Check grammar/punctuation/style (35; 0; 0; 0)
Consult APA publication manual (0; 0; 0; 4)
Get feedback on paper (29; 57; 30; 35)
Revise draft of paper (42; 57; 30; 38)
Copyright the research instruments (3; 0; 0; 0)
Submit paper for publication (19; 39; 52; 54)
Receive reviews (0; 11; 4; 4)
Make postreview revisions (0; 46; 22; 19)
Resubmit paper (0; 14; 13; 8)
Report final results to subjects (0; 0; 0; 8)
Conduct follow-up study (0; 0; 9; 0)
Receive notification of acceptance (0; 11; 4; 0)
Publish the research paper (anchor)

Copyright of Current Psychology is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.