

A cognitive analysis of number-series problems: Sources of individual differences in performance

JO-ANNE LEFEVRE and JEFFREY BISANZ
University of Alberta, Edmonton, Alberta, Canada

Two experiments were conducted to identify the roles of three hypothesized procedures in the solution of simple number-series problems and to determine whether individual differences in these solution procedures are related to performance on a number-series subtest from a standardized test of intelligence. The three procedures are recognition of memorized series, calculation, and checking. Subjects verified whether number sequences formed rule-based series. True series included both memorized counting series (e.g., "5 10 15 20") and unfamiliar noncounting series (e.g., "1 4 7 10"). False series could not be described by simple rules. The results of Experiment 1 indicated that (1) counting series were verified more quickly than were noncounting series, and (2) partial counting information in false series facilitated rejection. In Experiment 2, reliable differences in recognition of memorized sequences and calculational efficiency were found between individuals who scored well on a standardized test of number-series completion and those who scored poorly. The results provide a basis for understanding how individual differences in knowledge influence performance on problems often used to assess inductive reasoning skill.

Inductive reasoning, defined as inferring a general rule or relation from specific elements, is generally considered to be an integral component of many cognitive activities (Greeno, 1978; Holyoak, 1984; Pellegrino & Glaser, 1980, 1982; Simon & Lea, 1974). Reliable individual differences in inductive reasoning exist (Sternberg & Gardner, 1983), and psychometric tests of intelligence often consist in large part of items that are presumed to require inductive reasoning. Recent research on inductive reasoning has been focused on identifying the cognitive processes involved in psychometric tasks, such as analogy problems (Holzman, Pellegrino, & Glaser, 1982; Mulholland, Pellegrino, & Glaser, 1980; Sternberg, 1977) and series-completion problems (Holzman, Pellegrino, & Glaser, 1983). Another goal of recent research has been to identify the cognitive bases of individual differences in performance (Goldman & Pellegrino, 1984; Keating, 1984; Pellegrino & Glaser, 1979). The research reported in this paper is focused on the process of solving number-series completion problems and, specifically, on identification of the knowledge and processing components that contribute to differential problem-solving success.

Solution of series-completion problems is considered to be a prime example of inductive reasoning because a problem solver must detect or formulate a relation or rule

among elements in a series. Previous research on series completion has been based on a general model of solution that consists of four component processes (Holzman, Glaser, & Pellegrino, 1976; Holzman et al., 1983; Kotovsky & Simon, 1973; Simon & Kotovsky, 1963). Consider, for example, the problem "3 4 5 6 13 14 15 16 23 _ _," which consists of three, four-element periods. The first solution component, *detection of relations*, involves examination of the series and generation of a hypothesis about the relation that exists among adjacent elements (e.g., the relation is +1 among the first four elements). Relations among numbers may vary considerably in the type of arithmetic operation (e.g., addition, exponentiation) and the magnitude of the operation (e.g., +1 vs. +8). The second solution component, *discovery of periodicity*, involves detection of period boundaries or structure, and the third component, *completion of pattern description*, involves generation of a rule that accounts for the position of all elements, both between and within periods. The fourth component, *extrapolation*, is the means by which numbers are generated to fill the blanks.

The first three components combine to produce a pattern description. Conclusions about performance on series-completion problems are based primarily on a presumed relationship between pattern descriptions and working memory: more complex pattern descriptions correspond with greater demands on working memory. In support of this hypothesized relationship, memory-load requirements appear to contribute considerably to the difficulty of alphabetic series (e.g., "aaabbbcc_"; Kotovsky & Simon, 1973) and number series (e.g., "3 7 11 15 19 23 _"; Holzman et al., 1983). If working-memory capacity is important for performance, then individual differences in capacity should be related to differences in solution accuracy. Holzman et al. (1983) found

The research reported in this paper was supported by grants from the Natural Sciences and Engineering Research Council of Canada and from the General Research Fund, University of Alberta. The data described in Experiment 1 were presented at the annual meeting of the Canadian Psychological Association in Ottawa, May 1984. The authors are grateful to Mark Ashcraft, Gay Bisanz, Peter Dixon, Chris Herdman, and Robert Sternberg for their helpful comments, and to Cindy Scott for her assistance in several phases of the research. Requests for reprints should be sent to Jo-Anne LeFevre, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

qualified evidence with children in grades four and five for a relationship between individual differences in working-memory capacity and performance on number series, but no such evidence was found with adults.¹ Other factors, such as the type of arithmetic operation and the magnitude of arithmetic operations within the number series, also contributed to item difficulty. Holzman et al. also found that individual differences in performance on number series might be related to differences in computational skill among children.

These findings provide a preliminary basis for understanding performance on series-completion problems, but the four separate components of the model have not been well specified. A more detailed analysis of the component processes might be useful for explicating the nature of reasoning on this task, as well as providing a basis for detecting individual differences that are important for reasoning. The first component, detection of relations, is of special interest because it is an elementary case of rule induction and because it is involved, in some form, in the operation of the second and third components as well. In our research, we sought to provide a more detailed account of the processes involved in this critical first component and to determine whether individual differences in this component are related to scores on a psychometric test of reasoning that consists of a diverse set of number series. We hypothesize that three procedures are important for detecting relations among numbers: recognition of memorized numerical series, calculation, and checking.

Recognition of numerical series refers to a direct and relatively rapid comparison between a sequence of numbers presented visually and a corresponding sequence available in semantic memory. In existing formulations (Holzman et al., 1983; Kotovsky & Simon, 1973), no distinction is made between series such as "3 6 9 _" and "2 5 8 _": presumably, the same rule is induced in both cases and, therefore, processing demands are equivalent. In contrast, we hypothesize that the two problems are solved quite differently. We assume that familiar sequences of numbers, such as "3 6 9 12," are stored as units in semantic memory and can be retrieved relatively rapidly, much like sections of the alphabet (Klahr, Chase, & Lovelace, 1983). Thus, identification of a sequence in a number-series problem occurs when the memorial representation of that sequence is activated. Subsequent numbers in the sequence are generated by means of retrieval processes.

When processing unfamiliar sequences, such as "2 5 8 _," attempts to retrieve the sequence fail and *calculation* is invoked: Subjects calculate interelement differences (+3, +3), identify the rule that characterizes these differences, and then mentally add an appropriate number to the last element to obtain an answer. Thus, solution of problems such as "3 6 9 _" is assumed to be based primarily on recognition of familiar, memorized numer-

ical series, whereas solution of problems such as "2 5 8 _" requires calculation of interelement differences.

It is important to note that the distinction between "familiar" and "unfamiliar" sequences is entirely heuristic at this point. Our assumption that some sequences are stored as units in semantic memory is consistent with the observation that children commonly develop counting strings (e.g., "5 10 15 20...") that can be recited quickly and accurately (Ashcraft, Fierman, & Bartolotta, 1984). To operationalize the distinction between familiar and unfamiliar sequences, we assume initially that familiar sequences include common, memorized counting strings and that unfamiliar sequences include other strings that also have constant interelement differences.

These two procedures, recognition of memorized sequences and calculation, operate until a discrepant element is discovered, such as the fourth element in "3 6 9 13..." or "2 5 8 12..." These particular elements are discrepant in the sense that they violate the simple rule (+3) that relates the first three elements. A discrepant element may mark the boundary between two periods, and so detection of the discrepancy is important for initializing recognition of memorized sequences and calculation procedures on the new period. In problems without a periodic structure, detection of a discrepant element may indicate that the current rule is incorrect. At this point, the subject may reprocess the sequence to determine whether the discrepancy is simply the result of an error in the original procedure (e.g., an incorrect encoding or a mistaken calculation). This reprocessing is referred to as *checking*. We assume that checking is a relatively lengthy procedure that is invoked as a function of certain problem characteristics, much like the justification process described by Sternberg (1977). In particular, checking should be more evident when the discrepant element differs only slightly from the expected value (e.g., "2 5 8 11 13") than when it is completely anomalous (e.g., "2 5 8 11 93"). In the former case, the discrepancy is more likely to be due to a simple error in encoding or calculation, and thus checking would be appropriate for detecting the error; in the latter case, the discrepancy is likely to be too large to be the result of a simple encoding or calculation error.

We conducted two experiments, one to identify the roles of series recognition, calculation, and checking in the solution of simple number-series problems (Experiment 1), and the other to determine whether individual differences in these solution procedures are related to performance on a number-series subtest from a standardized test of intelligence (Experiment 2). In both experiments, subjects judged whether or not sequences of numbers formed "valid" series, that is, series that could be described by simple rules. This type of two-choice verification format has been employed to examine solution processes used on other tasks of inductive reasoning (e.g., Mulholland et al., 1980). It is especially appropriate in the present

context because our primary concern is with processes that contribute to detection of relations, rather than with subsequent processes (e.g., extrapolation).

EXPERIMENT 1

Adults were presented with four-element, numerical sequences and asked to respond to each one by pressing a "true" or a "false" key, depending on whether the sequence formed a valid number series. Valid series included both counting series (familiar number series) and noncounting series (unfamiliar series). If the distinction between counting series and noncounting series is plausible, then latencies for true counting series should be considerably faster than for true noncounting series.

Invalid and anomalous sequences were identical to true problems, except that one number was altered. For invalid problems, the altered number was only slightly different in magnitude from its correct counterpart, whereas for anomalous problems, the difference was very large. If checking is inversely related to the degree of discrepancy, then anomalous problems should be rejected much more quickly than invalid problems. Moreover, latencies for both types of problems can be examined as a function of the position of the discrepant element to determine whether solution procedures are self-terminating or exhaustive. Self-termination is often an efficient and characteristic mode of processing on problems of inductive reasoning (Mulholland et al., 1980; Sternberg, 1977). For true problems, processing must be exhaustive because all elements must be evaluated to determine whether they fit a rule. For invalid and anomalous problems, however, processing could terminate as soon as a discrepant element in the series is detected. If checking is not invoked and if self-terminating procedures are used with number series, then latencies for invalid and anomalous problems should be faster than for corresponding true problems, because solution of the latter requires exhaustive processing. If solution procedures operate from left to right, then latencies for invalid and anomalous problems also should increase monotonically as a function of the position of the discrepant element.

Method

Subjects. The sample included 16 university undergraduates, 8 males and 8 females. Two additional students were tested, but their data were discarded because of unusually high error rates (greater than 30%) on specific types of problems.

Materials. The six classes of problems presented in Table 1 were defined by the combination of series (counting, noncounting) and validity (true, invalid, anomalous). True counting series consisted of familiar sequences of four numbers. Each true counting series listed in Table 1 was presented three times, for a total of 18 problems. Both invalid counting series and anomalous counting series were derivatives of true counting series. Invalid counting series were constructed by adding 1 to, or subtracting 1 from, a number in a corresponding true counting series. This discrepant alteration appeared three times in each of the second, third, and fourth positions, for a total of 9 such problems. Similarly, 9 anomalous counting series were constructed by substituting a very large number in the

Table 1
Examples of Number-Series Problems Used in Experiment 1

Validity	Series Type		
	Counting Series	Noncounting Series	
True	1 2 3 4	2 5 8 11	
	2 4 6 8	1 4 7 10	
	1 3 5 7	4 7 10 13	
	3 6 9 12	1 5 9 13	
	4 8 12 16	2 6 10 14	
	5 10 15 20	3 7 11 15	
Invalid			
	Position 2	2 3 6 8	1 5 7 10
	Position 3	5 10 16 20	3 7 12 15
Position 4	4 8 12 15	2 6 10 15	
Anomalous			
	Position 2	5 48 15 20	3 45 11 15
	Position 3	1 2 31 4	1 4 36 10
Position 4	3 6 9 56	3 7 11 54	

Note—Discrepant elements are underscored.

second, third, or fourth positions of corresponding true counting series.

True noncounting series consisted of less familiar number sequences. Each of the true noncounting series in Table 1 was presented three times. Corresponding invalid noncounting and anomalous noncounting series were constructed in the same manner as were counting series.

Thus, the entire set consisted of 36 true, 18 invalid, and 18 anomalous problems. Two different presentation orders were constructed so that each third of the problem set (24 problems) contained (1) each of the 6 true counting series and 6 true noncounting series and (2) 3 of each of the four remaining classes of problems. The order of problems within each third of the set was randomized with the constraint that no more than 3 true series or 3 false series appeared consecutively.

Procedure. Problems were presented on a video display unit, and subjects responded by pressing an appropriate key on a response panel that contained three keys arranged horizontally. Subjects were instructed to place the index finger of the preferred hand below the middle key and to return the finger to that position after each keypress. Presentation of each trial was initiated by pressing the middle key. A string of four numbers was presented after a 700-msec delay, and the subject used the right and left keys to indicate whether the string represented a valid number series. Response latency and accuracy were recorded by a microcomputer. Following a response, the stimulus was removed from the screen and replaced by a prompt, which remained until the middle key was pressed again. Both position of the valid or "true" key and stimulus order were counterbalanced across subjects. When presented at a distance of .6 m from the subject, the series subtended 3.0° to 5.4° horizontally and 0.5° vertically.

Subjects were informed that a sequence of numbers forms a true number series if "the numbers are related to each other in a regular and systematic way," "they form a pattern," or if the sequence is described by a "regular pattern or rule." Examples of true and false series were provided. Subjects were also given 10 practice trials with feedback, including items from each of the six problem types. They were asked to answer as quickly as possible without making mistakes.

Results and Discussion

Accuracy was very high, averaging 98.6% overall and ranging from 93.7% (true noncounting series) to 100% (anomalous counting series). Median response latencies

for correct responses were computed, and a 2 (series: counting vs. noncounting) \times 3 (validity: true vs. invalid vs. anomalous) analysis of variance was performed with repeated measures on both factors. No effects or interactions involving sex were found in preliminary analyses, so this variable was omitted from subsequent analyses. Mean values are illustrated in Figure 1. Counting series were solved more rapidly than were noncounting series [$F(1,15) = 46.45, p < .001$]. Latencies also varied reliably by validity [$F(2,30) = 36.04, p < .001$] and the two factors interacted [$F(2,30) = 8.49, p < .01$]. Tests of simple effects and supplementary analyses of variance were employed to clarify these effects.

Latencies for true items. Latencies for true counting series were much faster than for true noncounting series [$F(1,45) = 41.90, p < .001$], in support of the distinction between these two series types. An alternative explanation for this effect is that some of the interelement differences for true counting series are smaller than the interelement differences for true noncounting series and that computation of larger differences simply takes longer (cf. Groen & Parkman, 1972). To test this hypothesis, a separate analysis was performed only on problems of both types for which the interelement differences were +3 and +4 (e.g., "3 6 9 12" for counting series and "2 5 8 11" for noncounting series). The +3 problems were

verified faster than were +4 problems [$F(1,15) = 8.77, p < .01$]. More importantly, true counting series were verified more quickly than were true noncounting series [$F(1,15) = 45.52, p < .001$], so the difference between the two types of series cannot be attributed solely to discrepancies in the size of interelement differences.

These results are consistent with the hypothesis that different processes operate for true counting series than for true noncounting series: Verification of counting series may be facilitated by means of counting-recognition processes, whereas noncounting series may be verified by means of slower processes that involve the calculation of interelement differences. Moreover, the difference between +3 and +4 problems indicates that all instances within the counting and noncounting categories are not homogeneous. These differences were explored further in Experiment 2.

Latencies for invalid items. For both counting series and noncounting series, latencies for invalid problems were longer than for true problems [$F(1,60) > 15, ps < .001$]. Because true problems appeared more frequently than invalid or anomalous problems, these latency differences might reflect an advantage of increased practice on true problems. However, even when latencies for the first third of the true problems were compared with latencies for all invalid and anomalous problems, true items were

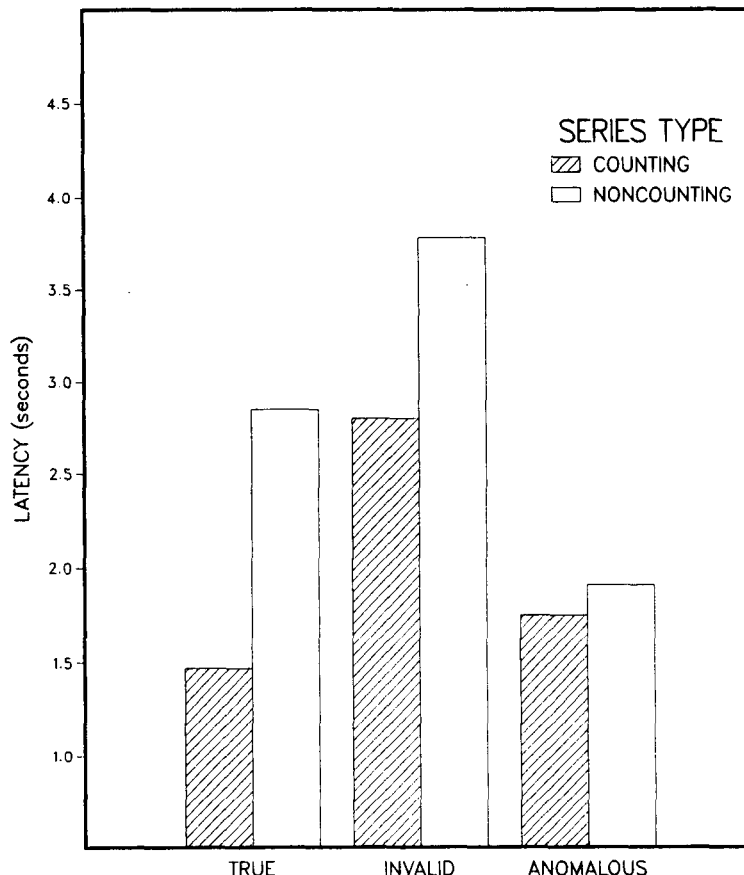


Figure 1. Mean verification latencies for the six problem types in Experiment 1.

solved more quickly. The fact that invalid noncounting series required considerably more solution time than true noncounting series is inconsistent with simple hypotheses about either self-terminating or exhaustive procedures that predict that no more than an equivalent amount of time would be spent on invalid as opposed to true problems. We suggest that, because invalid problems differ only slightly from their true counterparts, they may elicit a degree of checking to confirm that evaluation is correct. Such checking would be less likely for true problems because repeated calculations of the same interelement differences would partially confirm their accuracy.

Invalid noncounting series latencies were even slower than invalid counting series latencies [$F(1,15) = 19.06$, $p < .001$]. This difference may be attributable to facilitative effects of partial counting information that is present in invalid counting series but not in invalid noncounting series. More specifically, partial counting sequences among the first two or three elements of a series would make it possible for a subject to process the item rapidly, facilitated by counting recognition, up to the point of the discrepant element. The opportunity to use such partial counting sequences would exist only for invalid counting series and, if processes operate from left to right, the advantage would only be apparent when the discrepant element was in the rightmost positions. To test this hypothesis, an analysis was performed on median latencies for invalid counting series and invalid noncounting series problems with the discrepant element in the second, third, and fourth positions. The data are presented in the top portion of Figure 2 and are consistent with these conjectures. When the discrepant element was in the second position, the difference between counting series and noncounting series problems was negligible [$p > .25$]. In contrast, invalid counting series were solved more quickly than were invalid noncounting series when the discrepant

element was in the third or fourth positions [$ps < .001$], and latencies for invalid counting series decreased as a function of position [$F(1,60) = 6.88$, $p < .025$]. In effect, having the discrepant element in the second position made invalid counting series and noncounting series functionally equivalent by eliminating any facilitative effects due to recognition of partial counting sequences.

In summary, no evidence for self-terminating solution procedures was found for invalid problems. For invalid noncounting series, the relatively long solution latencies and the lack of a position effect are most indicative of exhaustive processing that includes checking. The descending function for invalid counting series was unexpected and is consistent with the notion that exhaustive procedures are facilitated by identification of partial counting information early in the solution process.

Latencies for anomalous items. For both counting series and noncounting series, latencies for anomalous problems were much shorter than for invalid problems [$Fs(1,60) > 20$, $ps < .001$] (see Figure 1), as would be expected if checking is involved only in the latter case. Anomalous counting series were solved somewhat faster than were anomalous noncounting series, but the difference was not reliable. The failure to find a reliable difference between anomalous counting series and noncounting series appears to indicate that the recognition procedure is not involved in solving anomalous problems. Such a conclusion is contradicted, however, by an analysis on median latencies for anomalous problems as a function of the position of the discrepant element. As illustrated in the bottom portion of Figure 2, the effect of position varied for counting series and noncounting series [$F(2,20) = 11.41$, $p < .001$]. For anomalous noncounting series, latencies increased linearly as a function of position [$F(1,60) = 17.32$, $p < .001$]. This result is consistent with the notion that processing proceeds from left to right and is terminated as soon as an anomaly is identified.

If processing is self-terminating for anomalous noncounting series, then the same type of processing might be expected for anomalous counting series, because the subject has no prior basis for distinguishing the two types. Consequently, an ascending function would be found for anomalous counting series. To the contrary, latencies for anomalous counting series declined somewhat as a function of position, but the effect was negligible. One possibility is that processing is exhaustive and partial counting information facilitates rapid analysis on these problems in much the same way as on invalid counting series. The need for checking is eliminated by the gross discrepancy of the anomalous element and, hence, latencies are much faster. The negative direction of the slope suggests that partial counting information is recognized more quickly when the leftmost elements of the counting string are not disrupted, as was the case for invalid counting series. The fact that the slope was not significant may indicate that a "floor" has been reached for recognition of four-element numerical series, an interpretation that is con-

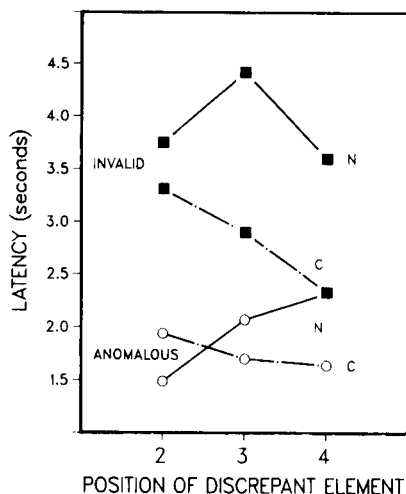


Figure 2. Mean verification latencies as a function of the position of the discrepant element for invalid and anomalous problems in Experiment 1. C refers to counting series; N refers to noncounting series.

sistent with the fact that latencies for anomalous counting series were similar to those for true counting series. A second possibility is that processing is self-terminating and facilitated by partial counting information when it is available. Self-termination would produce an ascending function, partial counting information would produce a descending function, and the combination could yield no overall effect of position. We favor the second explanation, because it suggests that anomalous counting series and noncounting series are both processed in a self-terminating fashion. In either case, partial counting sequences facilitate processing.

Thus, the data are consistent with the notion that verification of anomalous strings is facilitated by a relatively rapid process that involves neither computation of differences nor lengthy checking processes. Processing appears to be self-terminating, but for anomalous counting series, the usual position effects of self-termination are negated by facilitative effects of partial counting information.

Summary

Evaluation of simple number series involves a variety of procedures that are invoked selectively, depending on problem characteristics. The data are consistent with the hypotheses that (1) recognition of memorized sequences facilitates verification of certain numerical sequences (counting series) and (2) a self-terminating procedure permits relatively rapid rejection of some sequences containing grossly discrepant numbers (anomalous items). Verification of true and invalid noncounting series and invalid counting series depends on slower procedures that most likely involve mental calculations and comparisons of interelement differences. Solution procedures used on invalid problems appear to be exhaustive and to entail a degree of checking. Most striking was the facilitative effect of partial counting information on invalid counting series and, to a lesser extent, on anomalous counting series, especially when the discrepant element was in one of the rightmost positions.

EXPERIMENT 2

There were two major objectives of Experiment 2. The first was to look for individual differences in the procedures identified in Experiment 1. A large group of subjects were pretested with a number-series subtest from the Lorge-Thorndike Test of Intelligence (Lorge & Thorndike, 1957). Two smaller samples were isolated from either end of the range of scores. If expertise at solving number-series problems is related to individual differences in counting recognition (cf. Chi, Glaser, & Rees, 1982), then the high-scoring group might process more sequences in a uniformly fast manner than the low-scoring group. We also suspected that calculational efficiency might differ between high- and low-skill subjects. In sum, we were interested in determining whether relatively task-specific knowledge could contribute to differences in a measure of inductive-reasoning skill.

The second objective was to replicate and extend the counting/noncounting and position effects found in Experiment 1. We modified the problem set in a number of ways. First, the basic set of counting series was restricted to series with differences of 2, 3, 4, and 5, to make the average interelement differences equivalent between counting series and noncounting series. Second, the length of each series was extended from four elements to five, thus allowing a more stringent test of the various position effects found in Experiment 1. Finally, to determine whether a range of sequences that might be called counting series are in fact processed in a homogeneous fashion, two extra true counting series were added to the counting set, "1 2 3 4 5" and "6 12 18 24 30," and performance for each of the true counting series was analyzed separately.

Method

Pretest. One hundred eight undergraduates were tested in group sessions lasting about 35 min. Each person received course credit for participating. Three tests were administered: the number-series subtest from the Lorge-Thorndike Intelligence Test (Lorge & Thorndike, 1957); the Addition Test (N-1) from the French Kit (French, Ekstrom, & Price, 1963), which consists of three-term, two-digit addition problems; and the Subtraction and Multiplication Test (N-3), also from the French Kit, which consists of two-term subtraction and multiplication problems. The latter two, which are essentially speeded tests of fluency with simple arithmetic computations, were included in order to provide an independent test of simple arithmetic skill. The number-series subtest consists of 28 number-series-completion problems in which subjects select the correct answer from five alternatives. The number-series subtest was administered first with a 9-min time limit. The arithmetic test was administered next, each with the standard 4-min time limit. Speed and accuracy were emphasized in the instructions. No subject completed all of the items on any test. The mean number of items answered correctly on the number-series subtest, with 28 problems, was 13.8.

Subjects. Sixteen subjects were selected according to their scores on the Lorge-Thorndike number-series subtest. One subject was later replaced because of his high error rate on the experimental task. Five of these subjects were paid \$5 each for participating, and the others received course credit. Eight of these subjects (4 males and 4 females) were selected from the lower 26% of the total sample and were designated as "low skill"; the remaining 8 (4 males and 4 females) were selected from the upper 12% and were designated "high skill." Mean Lorge-Thorndike scores for the low- and high-skill groups were 8.6 and 20.5, respectively.

Materials and Procedure. The procedure was identical to that used in Experiment 1. Stimuli consisted of five-element number strings but were otherwise similar to those in Experiment 1. Each of the four true counting series and true noncounting series in Table 2 was presented eight times, for a total of 64 valid items. Another 64 false items (32 invalid and 32 anomalous) were constructed as in Experiment 1, except that the discrepant elements appeared in each of four positions an equal number of times. Examples are presented in Table 2. In addition to those 128 items, two extra true counting series—"1 2 3 4 5" and "6 12 18 24 30"—were each presented eight times, for a total of 144 items.

Two presentation orders were constructed such that, within each quarter of the problem set (32 problems), each true item appeared twice and the discrepant element appeared an equal number of times in each of the four positions for both invalid and anomalous items. The order of the problems was randomized with the constraint that no more than three true or false problems appeared in a row.

Table 2
Examples of Number-Series Problems Used in Experiment 2

Validity	Series Type	
	Counting Series	Noncounting Series
True	2 4 6 8 10	2 5 8 11 14
	3 6 9 12 15	4 7 10 13 16
	4 8 12 16 20	3 7 11 15 19
	5 10 15 20 25	5 9 13 17 21
Invalid		
	Position 2: 2 5 6 8 10	4 8 10 13 16
	Position 3: 3 6 8 12 15	2 5 9 11 14
	Position 4: 4 8 12 15 20	5 9 13 18 21
	Position 5: 5 10 15 20 24	3 7 11 15 20
Anomalous		
	Position 2: 5 8 15 20 25	3 8 11 15 19
	Position 3: 4 8 7 16 20	5 9 7 17 21
	Position 4: 3 6 9 6 15	2 5 8 6 14
	Position 5: 2 4 6 8 9 2	4 7 10 13 9 2

Note—Discrepant elements are underscored.

Results

The 16 extra true problems were omitted in the following analyses, except as noted. As in Experiment 1, accuracy was high, averaging 97.2% overall. Median response latencies were computed and analyzed in a 2 (skill: low vs. high) × 2 (sex: male vs. female) × 2 (series: counting vs. noncounting) × 3 (validity: true vs. invalid vs. anomalous) analysis of variance with repeated measures on the last two factors.² As in Experiment 1, counting series were solved more quickly than were noncounting series [$F(1,12) = 72.51, p < .001$], response times varied with validity [$F(2,24) = 28.88, p < .001$], and the two factors interacted [$F(2,24) = 5.46, p = .01$]. Furthermore, high-skill subjects were faster than low-skill subjects [$F(1,12) = 6.68, p < .05$].

Skill also interacted with series [$F(1,12) = 5.23, p < .05$], and the interaction of skill, series, and validity just missed the conventional level of significance [$F(2,24) = 3.07, p = .065$]. Mean values are illustrated in Figure 3. Tests of simple effects and supplementary analyses of variance were used to clarify these effects.

Latencies on true items. True counting series were solved more quickly than were true noncounting series [$F(1,36) = 62.52, p < .001$]. To identify possible differences among counting series and between groups, latencies on the four true counting series shown in Table 2, plus the two extra true counting series, were analyzed in a 2 (skill: low vs. high) × (2 sex: male vs. female) × 6 (counting series: +1, +2, +3, +4, +5, +6) analysis of variance with repeated measures on the last factor. High-skill subjects were faster than low-skill subjects [$F(1,12) = 6.10, p < .05$], response time varied with the type of counting series [$F(5,60) = 19.35, p < .001$], and these two factors interacted [$F(5,60) = 2.89, p < .05$]. Mean values are shown in Figure 4.

For the +1 (1 2 3 4 5), +2 (2 4 6 8 10), and +5 (5 10 15 20 25) counting series, there was no latency difference between high- and low-skill subjects [$F_s(1,72) < 1$]. In contrast, for the +3 (3 6 9 12 15) series, the skill difference approached significance [$F(1,72) = 3.36, p < .1$], and for the +4 (4 8 12 16 20) and +6 (6 12 18 24 30) series, high-skill subjects responded more quickly than low-skill subjects [$F_s(1,72) > 9, p_s < .01$]. Thus, there were reliable skill differences on a subset of the counting series, suggesting that all such series were not processed in the same way.

A related issue involves within-group counting series differences. Neuman-Keuls analyses of the verification times for the six counting series were done within each

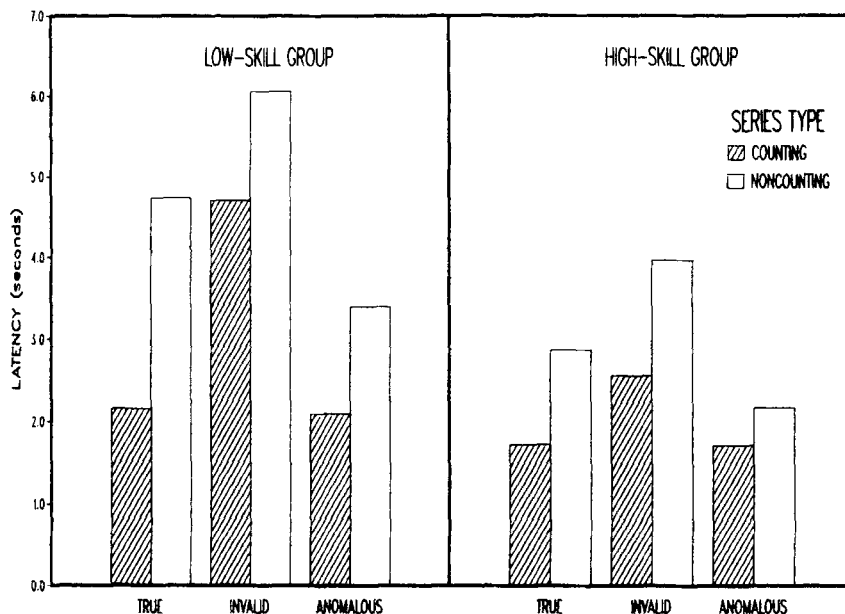


Figure 3. Mean verification latencies for low- and high-skill subjects on the six problem types in Experiment 2.

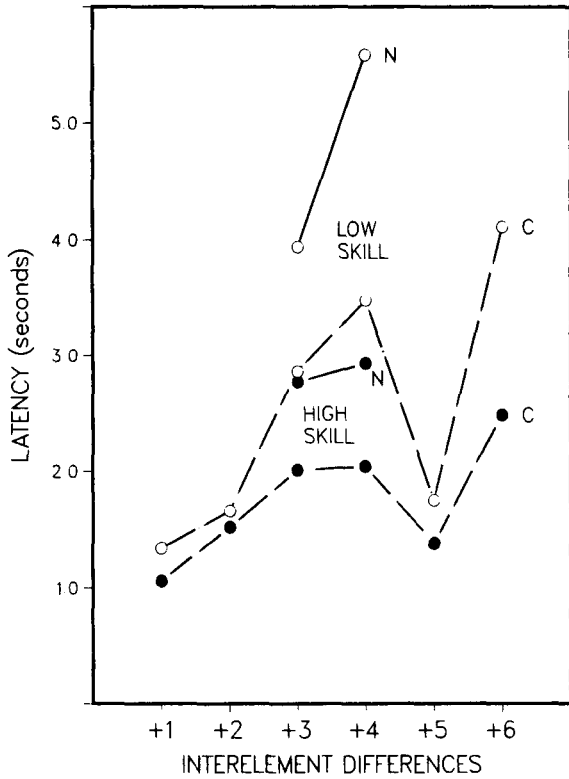


Figure 4. Mean verification latencies for true problems as a function of interelement differences in Experiment 2. C refers to counting series; N refers to noncounting series.

skill group, using an alpha level of .01. For low-skill subjects, the +1, +2, and +5 counting series did not differ from each other, although all were verified significantly faster than the +3, +4, and +6 series, and the +3 series was verified more quickly than the +6 series. For high-skill subjects, however, the only difference was that “6 12 18 24 36” was verified more slowly than “1 2 3 4

5.” Verification times across counting series for individual subjects are illustrated in Figure 5. Essentially, the variability across counting series and skill levels supports the conclusions drawn from the Neuman-Keuls analyses. All subjects were relatively fast and uniform on the +1, +2, and +5 counting series, but low-skill individuals were extremely variable on the other counting series. High-skill subjects were less variable across all counting series, with the possible exception of +6.

Another potential source of individual differences in number-series solution is calculational efficiency. Analysis of the true noncounting items (the short lines in Figure 4) suggests that speed of calculation is a source of individual differences on number-series problems. A 2 (skill) × 2 (sex) × 2 (size: +3 vs. +4) analysis of variance indicated that, in addition to main effects of skill and size, there was a reliable interaction of these two factors [$F(1,12) = 22.96, p < .001$]. Low-skill subjects verified +3 problems more quickly than +4 problems [$F(1,12) = 56.17, p < .001$], whereas high-skill subjects did not [$F < 1$]. Thus, not only are low-skill subjects slower at calculating interelement differences than high-skill subjects, but the size of the interelement difference only influences latencies for the low-skill subjects. This result suggests that low-skill subjects use less efficient mental calculation processes. Two additional observations support this conclusion. First, the differences between low- and high-skill subjects was relatively small on items that could be solved without calculations (435 msec on true counting series) but very large on items that require calculations (1,864 msec on true noncounting series). This discrepancy between groups may even be an underestimate of the difference in calculational efficiency if some low-skill subjects were using mental calculations on some true counting series. Second, high-skill subjects solved many more problems than low-skill subjects on the speeded tests of arithmetic fluency (122.8 vs. 76.1 cor-

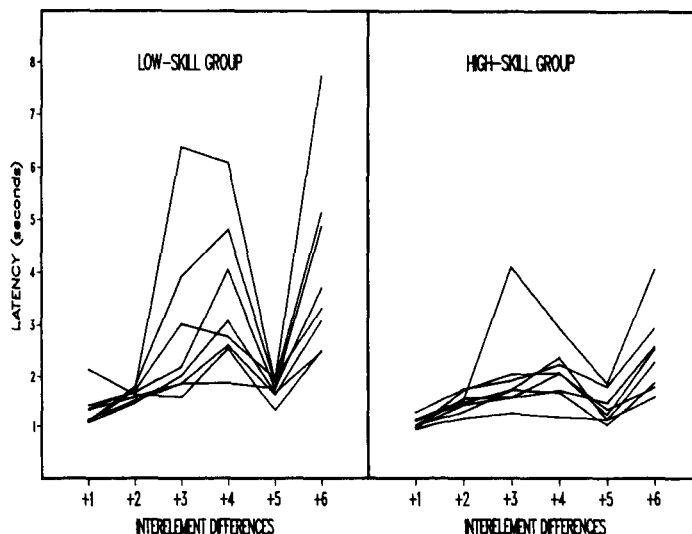


Figure 5. Median verification latencies for each subject on the six true counting series in Experiment 2.

rect across the two tests). More generally, the correlation between Lorge-Thorndike scores and scores on the arithmetic fluency tests for the total sample were moderate but significant [$r(106) = .36$ for the Addition Test, and $r(106) = .29$ for the Subtraction and Multiplication Test, $ps < .01$].

In sum, analyses of the true data indicate that (1) a substantial latency difference exists between counting series and noncounting series, (2) some numerical sequences are recognized quickly by all subjects, (3) high-skill subjects process different counting series in a more homogeneous fashion than do low-skill subjects, and (4) calculation slows the latencies of low-skill subjects more than those of high-skill subjects. Interestingly, although low-skill subjects may process fewer sequences at a very fast rate, they benefit from the facilitative effects of counting information more than do high-skill subjects because their calculations are so time-consuming.

Latencies on invalid items. As in Experiment 1, invalid problems were analyzed according to the position of the discrepant element in a 2 (skill: low vs. high) \times 2 (sex: male vs. female) \times 2 (series: counting vs. noncounting) \times 4 (position: 2 vs. 3 vs. 4 vs. 5) analysis of variance with repeated measures on the last two factors. High-skill subjects were marginally faster than low-skill subjects [$F(1,12) = 4.32, p = .06$]. Counting series were solved more quickly than noncounting series [$F(1,12) = 22.31, p < .001$], latency varied as a function of position [$F(3,36) = 4.90, p < .01$], and these two factors interacted [$F(3,36) = 3.80, p < .05$]. Mean values are shown in the top part of Figure 6.

As in Experiment 1, the facilitating effect of partial counting information was evidenced by the fact that verification times for invalid counting series varied as a function of position [$F(3,72) = 6.42, p < .001$], and the decreasing linear component of the trend was significant [$F(1,72) = 16.87, p < .01$]. In contrast, latencies for invalid noncounting series varied only marginally with position ($p < .10$).

In sum, the results of the position analysis for invalid problems are consistent with those found in Experiment 1 and support the conclusions that partial counting information facilitates solution, and checking processes contribute to relatively long latencies on these problems. In addition, this experiment demonstrated that partial counting information facilitates rejection latency on invalid items for both low- and high-skill subjects.

Latencies on anomalous items. As in Experiment 1, latencies on anomalous problems were much faster than on invalid problems for both counting series and noncounting series [$Fs(1,48) > 31, ps < .001$], true counting series and anomalous counting series did not differ [$F(1,48) < 1$], and anomalous noncounting series were solved more quickly than true noncounting series [$F(1,48) = 10.95, p < .01$]. Contrary to the results in Experiment 1,

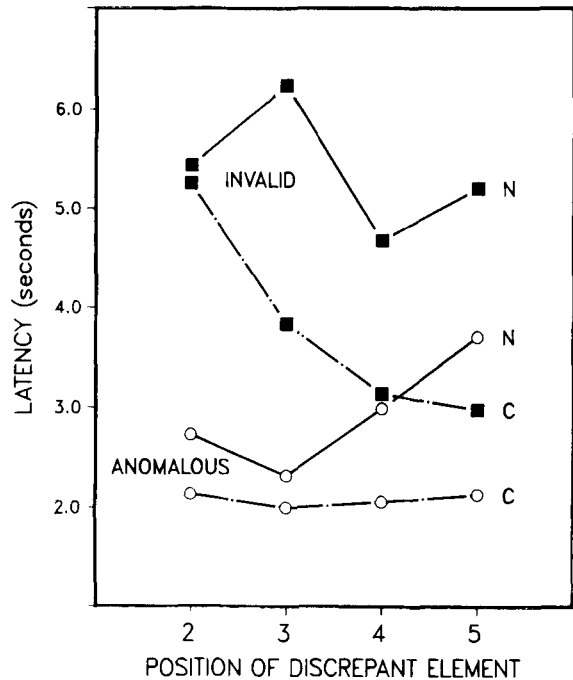


Figure 6. Mean verification latencies as a function of the position of the discrepant element for invalid and anomalous problems in Experiment 2. C refers to counting series; N refers to noncounting series.

anomalous counting series were rejected more quickly than anomalous noncounting series [$F(1,36) = 13.9, p < .01$].

Latencies for anomalous problems were analyzed in a 2 (skill: low vs. high) \times 2 (sex: male vs. female) \times 2 (series: counting vs. noncounting) \times 4 (position) analysis of variance with repeated measures on the last two factors. High-skill subjects were faster than low-skill subjects [$F(1,12) = 5.61, p < .05$], and counting series were solved more quickly than noncounting series [$F(1,12) = 17.62, p < .001$]. Although there was no overall effect of position, series interacted with position [$F(3,36) = 9.84, p < .001$] and these two factors interacted with skill [$F(3,36) = 5.08, p < .01$].

The interaction of series and position is shown in the bottom part of Figure 6. The configuration is very different from that for invalid problems. There was no effect of position for counting series, and an increase as a function of position for noncounting series occurred. Qualitatively, these results were similar to those obtained for the anomalous problems in Experiment 1 (compare Figure 2 and Figure 6). Differences between the two experiments appear to be due primarily to two factors. First, the average latency for noncounting series was greater than for counting series in Experiment 2, because longer series were used, and the mean latency for noncounting series was influenced more by the relatively lengthy times to reject items with the discrepant element in later positions.

Second, the performance of low-skill subjects differed markedly from that of both high-skill subjects and the subjects in Experiment 1 (see Figure 7).

For both high- and low-skill subjects, latencies did not vary as a function of position on anomalous counting series, a finding that replicates the results of Experiment 1. The two groups differed considerably in their performance on noncounting series, however. For high-skill individuals, the linear component of the position effect was significant [$F(1, 144) = 3.97, p < .05$], representing an increase of about 300 msec per position. This moderate increase is comparable to that found in Experiment 1 (approximately 400 msec per position). For low-skill individuals, the effect of position is much more striking. The elevation of the Position 2 latency was due primarily to one subject who was extremely slow on certain problems. When his latencies are excluded, latencies for noncounting problems increase linearly as a function of position. However, this increase is about 1 sec per position, more than three times that of the increase for high-skill subjects, and over twice that of the Experiment 1 subjects.

The results with anomalous items are consistent with our interpretations of performance in Experiment 1, although some modification is required to account for the data of low-skill individuals. Processing of anomalous items is relatively rapid and self-terminating, although for counting series, the facilitative effects of counting information balance the position effects usually associated with self-termination. For noncounting series, the magnitude of the slope for low-skill subjects in Figure 7 indicates that they invoke slow calculational processes from the beginning, whereas high-skill subjects either detect anomalies without calculating or calculate very quickly. Counting series information clearly benefits low-skill subjects more than high-skill subjects.

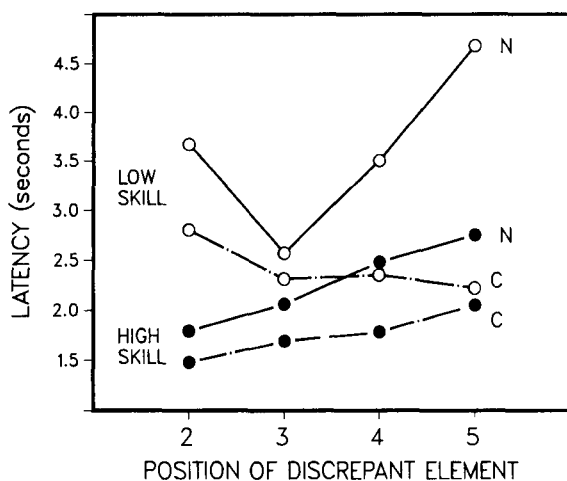


Figure 7. Mean verification latencies as a function of the position of the discrepant element for anomalous problems in Experiment 2. C refers to counting series; N refers to noncounting series.

Summary

The results of this experiment generally replicated those of Experiment 1. In addition, this experiment provided evidence for sources of individual differences in solution of number-series problems. First, high- and low-skill subjects were equally fast on some counting series (+1, +2, and +5), but low-skill subjects were generally slower and more variable on other counting series (+3, +4, +6). Second, partial counting information facilitated performance of high- and low-skill subjects to a similar degree on invalid problems, but low-skill subjects benefited more on anomalous problems. Low-skill subjects were slower at calculation and, hence, were more likely to benefit in situations where recognition of memorized sequences could be used. The fact that the two groups did not show different degrees of benefit on invalid items may indicate that high-skill subjects are more likely to check their calculations than are low-skill subjects.

GENERAL DISCUSSION

Previous models of series-completion solution (Holzman et al., 1983; Kotovsky & Simon, 1973) include four processes: relations detection, discovery of periodicity, completion of pattern description, and extrapolation. In this paper, we focused on relations detection in an attempt to specify the component procedures that might be involved in the initial stages of formulating a pattern description or rule. At least three procedures are involved in relations detection: recognition of memorized sequences, calculation, and checking. These three procedures were utilized in various combinations to verify simple number series. Recognition was used to quickly verify counting, but not noncounting, sequences. Similarly, partial counting information facilitated processing of the sequence and hence rejection of both invalid and anomalous counting series. Calculation of interelement differences was used when a number sequence could not be retrieved from memory. Checking, or partial reprocessing, was used on invalid problems that differed only slightly from valid series. Finally, these procedures were self-terminating on problems that contained highly anomalous elements and were obviously false, thus saving unnecessary processing. Individual differences in the component procedures were also identified: High-skill subjects used recognition of memorized series on a wider range of number sequences and calculated more efficiently than did low-skill subjects.

When several procedures are involved in a routine such as relations detection, the question that arises is how these procedures are organized and implemented. Siegler and Shrager (1984) described a model of how children solve simple arithmetic problems (e.g., $2 + 3$) that may also be applicable to relations detection in our task. In their model, an initial fast retrieval of the correct answer is attempted. If retrieval is unsuccessful, then slower, more elabora-

tive processes (e.g., counting) are used to find an answer. In our task, subjects first attempted to match the presented number sequence to a memorized series. If the attempt to match the series was unsuccessful, they then used slower, more effortful processes (e.g., calculation) to assist verification. In our studies, certain problems (e.g., "5 10 15 20 25") reliably elicited recognition of memorized sequences in all subjects. Other problems (e.g., "4 8 12 16 20") elicited recognition of memorized sequences for some subjects and calculation for others. These results suggest that the particular procedures used during relations detection depend both on the characteristics of the problem and on the knowledge base of the individual solving the problem. Knowledge, in this case, is indexed by accurate and rapid retrieval, rather than by accuracy alone.

Series-completion problems are often used as a measure of inductive reasoning, so we can speculate as to whether and how the differences between high- and low-skill subjects in Experiment 2 are related more generally to individual differences in reasoning. The differences in recognition of memorized sequences and calculation between high- and low-skill individuals are consistent with two recent observations about performance on various inductive reasoning tasks: (1) estimates of various information-processing parameters correlate significantly with factor loadings that represent psychometric measures of inductive reasoning (Sternberg & Gardner, 1983) and (2) adults who score well on tests of inductive reasoning generally process information more efficiently than other adults (Goldman & Pellegrino, 1984).

Our findings provide some basis for refining these generalizations. High- and low-skill subjects differed from each other on both forms of rule induction that were assessed: a rapid process (recognition of memorized sequences) and a slower, more elaborative process (calculation). The nature of the difference varied, however: In the case of recognition of memorized sequences, low-skill subjects were as efficient as high-skill subjects on some stimuli, but on others they were variable and generally less efficient; in the case of calculation, low-skill subjects were consistently less efficient. Both types of skill appear to be relatively domain-specific. The ability to recognize patterns or configurations of information in the domain is a characteristic of expertise in chess (Chase & Simon, 1973), electronics (Egan & Schwartz, 1979), physics (Chi et al., 1982), and even baseball (Chiesi, Spilich, & Voss, 1979). Similarly, striking individual differences exist in manipulating numerical information (e.g., Hunter, 1977). Recognition of memorized series and calculation may well be *domain-specific* forms of knowledge, but they probably are not *task-specific*: both procedures may be applicable to a wide variety of tasks that require numerical skills (Holzman et al., 1982; Muth, 1984). In contrast, checking and self-termination are related to efficiency and may well be both *task-* and *domain-independent*, but neither characteristic appeared to be related to skill in Experiment 2. Thus, we speculate that differences in efficiency

between high- and low-skill individuals may be most evident for processes that are relatively domain-specific.

In the present studies, we examined an aspect of performance on a particular task, rather than seeking to identify processes that apply to many tasks. This approach is quite limited in scope, but it does have some advantages. Consider, for example, the research by Sternberg and Gardner (1983) on inductive reasoning. Sternberg and Gardner postulated such components as *inference*, *mapping*, and *justification* to account for performance on series-completion problems and on other tasks of inductive reasoning. These postulated components appear to have some degree of commonality across different tasks, but as Goldman and Pellegrino (1984, p. 194) noted, the processes to which these labels refer are not well described (see also Pellegrino & Lyon, 1979). Indeed, the task analysis provided by Sternberg and Gardner is not sufficient for distinguishing such vastly different procedures as recognition of memorized sequences and calculation, which clearly need to be incorporated into any comprehensive model of number-series solution. Thus, the present approach, although limited in its initial generality, complements Sternberg and Gardner's approach by providing a more finely grained view of underlying solution procedures and, hence, a more detailed perspective on potential sources of individual differences.

REFERENCES

- ASHCRAFT, M. H., FIERMAN, B. A., & BARTOLOTTA, R. (1984). The production and verification tasks in mental addition. *Developmental Review*, *4*, 157-170.
- BRAINERD, C. J., & KINGMA, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, *17*, 210-247.
- CHASE, W. G., & SIMON, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.
- CHI, M. T. H., GLASER, R., & REES, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-76). Hillsdale, NJ: Erlbaum.
- CHIESI, H., SPILICH, G. J., & VOSS, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, *18*, 257-273.
- EGAN, D. E., & SCHWARTZ, B. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, *7*, 149-158.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, I. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.
- GOLDMAN, S. R., & PELLEGRINO, J. W. (1984). Deductions about induction: Analyses of developmental and individual differences. In R. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 2, pp. 149-198). Hillsdale, NJ: Erlbaum.
- GREENO, J. G. (1978). Natures of cognitive abilities. In W. K. Estes (Ed.), *Handbook of learning and cognition* (Vol. 5, pp. 239-270). Hillsdale, NJ: Erlbaum.
- GROEN, G. J., & PARKMAN, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, *79*, 329-343.
- HOLYOAK, K. (1984). Analogical thinking and human intelligence. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 199-230). Hillsdale, NJ: Erlbaum.
- HOLZMAN, T. G., GLASER, R., & PELLEGRINO, J. W. (1976). Process training derived from computer simulation theory. *Memory & Cognition*, *4*, 349-356.

- HOLZMAN, T. G., PELLEGRINO, J. W., & GLASER, R. (1982). Cognitive dimensions of numerical rule induction. *Journal of Educational Psychology*, **74**, 360-373.
- HOLZMAN, T. G., PELLEGRINO, J. W., & GLASER, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology*, **75**, 603-618.
- HUNTER, I. M. L. (1977). Mental calculation. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press.
- KEATING, D. P. (1984). The emperor's new clothes: The "new look" in intelligence research. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 1-46). Hillsdale, NJ: Erlbaum.
- KLAHR, D., CHASE, W. G., & LOVELACE, E. A. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **9**, 462-477.
- KLAPP, S. T., MARSHBURN, E. A., & LESTER, P. L. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, **112**, 240-264.
- KOTOVSKY, K., & SIMON, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology*, **4**, 399-424.
- LORGE, I., & THORNDIKE, R. L. (1957). *The Lorge-Thorndike intelligence test, levels A-H*. Boston: Houghton-Mifflin.
- MULHOLLAND, T. M., PELLEGRINO, J. W., & GLASER, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, **12**, 252-284.
- MUTH, K. D. (1984). Solving arithmetic word problems: Role of reading and computational skill. *Journal of Educational Psychology*, **76**, 205-210.
- PELLEGRINO, J. W., & GLASER, R. (1979). Cognitive components and correlates in the analysis of individual differences. *Intelligence*, **3**, 187-214.
- PELLEGRINO, J. W., & GLASER, R. (1980). Components of inductive reasoning. In R. E. Snow, P. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction. Vol. 1: Cognitive process analyses of aptitude* (pp. 177-217). Hillsdale, NJ: Erlbaum.
- PELLEGRINO, J. W., & GLASER, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 269-345). Hillsdale, NJ: Erlbaum.
- PELLEGRINO, J. W., & LYON, D. R. (1979). The components of a componential analysis. *Intelligence*, **3**, 169-186.
- SIEGLER, R. S., & SHRAGER, J. (1984). Strategy choices in addition: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills*. Hillsdale, NJ: Erlbaum.
- SIMON, H. A., & KOTOVSKY, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, **70**, 534-546.
- SIMON, H. A., & LEA, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Hillsdale, NJ: Erlbaum.
- STERNBERG, R. J. (1977). *Intelligence, information-processing and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.
- STERNBERG, R. J., & GARDNER, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General*, **112**, 80-116.

NOTES

1. Holzman et al. (1983) used backward digit span as a measure of working-memory capacity. Given more recent evidence that working memory and short-term memory may be functionally distinct (Brainerd & Kingma, 1985; Klapp, Marshburn, & Lester, 1983) and that backward digit span may reflect short-term rather than working memory, failure to find compelling correlations between backward digit span and performance on number series may not be a satisfactory basis for rejecting the hypothesized relationship.

2. There were no main effects of sex in the response time analyses. Sex was involved in some interactions, but the failure to find similar effects in Experiment 1 limits the generality of the results. Information about effects related to sex is available on request.

(Manuscript received September 30, 1985;
revision accepted for publication February 8, 1986.)