DOCUMENT RESUME

ED 412 605 EA 028 656

AUTHOR Allison, Derek J.; Morfitt, Grace
TITLE Problem Processing and the Principalship: Design, Methods and Procedures.
SPONS AGENCY Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).
PUB DATE 1996-04-00
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996). For a related paper, see EA 028 655.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Administrator Responsibility; Cognitive Ability; *Cognitive Measurement; Cognitive Processes; *Data Collection; *Educational Experience; Elementary Education; Foreign Countries; *Principals; Problem Solving; Productive Thinking; *Schemata (Cognition)
IDENTIFIERS *Cognitive Approaches to School Leadership Project

ABSTRACT

The Cognitive Approaches to School Leadership (CASL) Project builds on cognitive-science theories of how people think about problems, with special reference to related studies of administrative practice in general and educational administration in particular. This paper is the second of two papers that reports findings from clinical studies of expertise in the elementary principalship. The study's main objective was to analyze how individuals with greater administration expertise and individuals with lesser expertise in school leadership think through domain problems. The paper describes the theory base and provides a detailed account of the study design. A total of 31 elementary principals and 25 novice subjects participated in the study. Data-collection methods included six pencil-and-paper measures of personal variables, think-aloud responses to a case problem, and a juried review of the think-aloud processes. The paper describes the reputational and performance measures used by the jury to assign principals to expertise categories, and explains how the imputed-expertise categories were constructed. Three tables and two figures are included. (Contains 31 references.) (LMI)

Problem processing and the principalship:

Design, methods and procedures.

Derek J. Allison & Grace Morfitt
The University of Western Ontario

Revised version of a paper presented as part of the symposium
Clinical assessments of practical performance in school leadership: Findings from novice-expert studies of the elementary principalship.
Session #41.22 at the Annual Meeting of the
American Educational Research Association
New York City, April 1996

The research reported in this paper was supported by the Social Sciences and Humanities Research Council of Canada under grant #410-92-0329.

CASL Project Paper #AERA96 2.2, revised July 1996.

Design, methods and procedures

Derek J. Allison & Grace Morfitt

This paper is the second of two prepared to introduce a 1996 symposium describing and reporting findings from clinical studies of expertise in the elementary principalship conducted as part of our CASL Project. The first introductory paper offered an overview of the cognitive science theory of problem processing expertise and the studies of novice-expert differences underlying the research, together with a discussion of two major design issues, namely the use of think-aloud protocols and the problem of identifying novices and experts. This second paper provides a detailed account of the design of the study yielding the data discussed in the three analytical papers prepared for the symposium. The paper begins with a brief recapitulation of the theory base before moving to consider the details of the design.

Theoretical frame

Our Cognitive Approaches to School Leadership Project builds on cognitive science theories of how people think about problems (e.g. Reiman & Chi, 1989; VanLehn, 1989) with special reference to related studies of administrative practice in general (e.g. Cowan, 1990; Srivastva, 1983) and educational administration in particular (Hallinger, Leithwood & Murphy, 1993; Leithwood & Steinbach, 1995). Clinical studies of how elementary principals and others view and think about problems form a major line of inquiry in the CASL Project. Descriptions and some of the results from the first of these principal problem processing [PPP] studies were reported by Allison and Nagy (1991), and Allison and Allison (1993). This paper describes the design and conduct of our second PPP study. Both studies assumed that expertise in the elementary school principalship—as in previously studied specialized domains of human action—rests on the development and accurate access of task relevant declarative and procedural knowledge stored in schematically organized memory. This richly interconnected domain knowledge should enable principals to form more accurate and representative understandings of the tasks, problems and challenges with which they must deal than can novices who—by definition—have had little direct exposure to the domain. A similar difference is considered to hold within the population of school principals: those who have developed greater expertise than their colleagues should be able to provide fluid, practised responses to commonly encountered situations, and to draw on a richer repertoire of

CASL Project Paper #AERA96 2.2, revised July 1996.

strategies and possible solution ingredients when dealing with less frequently encountered problems which they see as more ill-structured. Think-aloud protocols of how individuals with varying levels of expertise work on a presented problem situation characteristic of the school principalship would therefore be expected to reflect these and related differences, while also capturing samples of problem related schematic memory which those with greater expertise bring to bear on the problem. Our main objective in the PPP studies has thus been to compare and analyze how individuals with greater and lesser expertise in school leadership think about and through domain problems with a view to (1) relating differences to previous studies of domain expertise and (2) identifying problem processing patterns and ingredients associated with expertise in the principalship. We are also interested in, and have been collecting data pertinent to, relationships between indicators of domain expertise and personal variables that may mediate the development of the schematic memory that is theorized as underlying expertise.

Conceptual design

Figure 1 provides a conceptual overview of the design adopted for our second PPP study. This Figure contains three conceptual sub-assemblies which will be discussed in turn: (1) a model of the task environment of principals, contained in the circle in the upper right corner of the Figure; (2) a model of the development of the domain relevant schematic knowledge considered to underlie expertise, located in the left two-thirds of the Figure, and; (3) flow-chart paths showing the derivation of datasets to be considered in the analysis, which culminate in the lower right of the Figure.

Task environment

As understood in our conceptual framework, the task environment of a school principal equates to a specific position as officially defined and informally constituted within a formal, intendedly bureaucratic (in the technical sense), organization. In essence, the work environment consists of all of the tasks, goals and responsibilities that principal X of school Y in educational system Z is expected to complete and assume, all of the specific problems and projects which she finds, is assigned, or otherwise becomes aware of which are associated with the successful and timely accomplishment of her tasks and responsibilities, all of the hopes, dreams and desires which she and others attach to the job. This environment is defined in part by the tasks, expectations, problems, values and standards associated with generic professional, theoretical and societal understandings of the

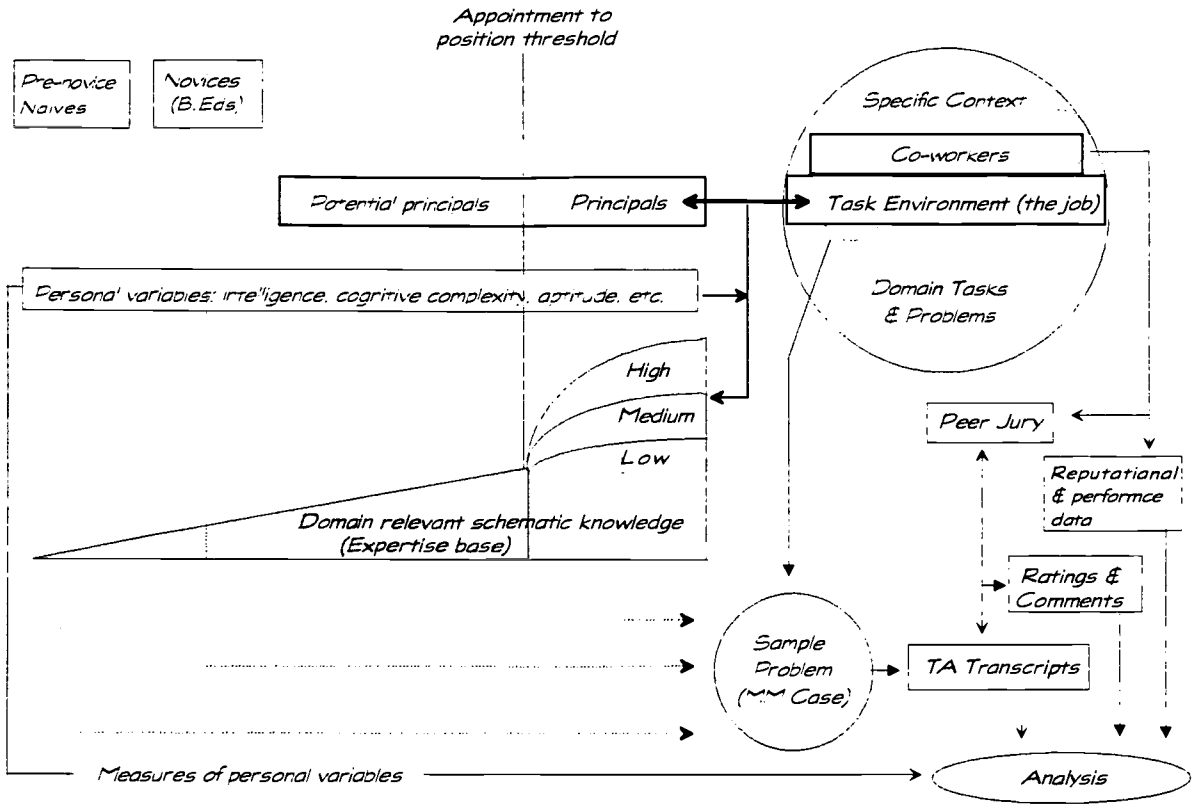

Figure 1
Conceptual design of second CASL PPP study.

principal's role and the work to be done, as represented by the lower half of the circle. These will be modified, extended, and circumscribed by specific contextual elements associated with a given position, as manifest in applicable formal documents, entrenched conventions, school history, traditions, culture, and the circumstances of the communities from which the school draws its pupils and other participants. Centrally important among the specific contextual factors will be the co-workers—teachers, support staff, senior administrators—with and through whom principals must work. The task environment exists regardless of the person occupying the specific principalship thus defined, although it is not independent of the role incumbent, and his or her actions, values, successes, or failures, which will modify aspects of the task environment through interaction.

Development of expertise

The main elements of Figure 1 that seek to model the development of administrative expertise are the principal box, located in the upper centre of the Figure, the model of domain expertise placed toward the centre of the Figure, and the personal variables box shown as intervening between these two elements.

Role incumbents. In the design sketched in Figure 1, the specific person occupying a given principalship at a given time resides (conceptually) in the oblong box linked to the task environment by the heavy, double arrowed, line. The person who is the role incumbent is portrayed as separate from the position itself to recognize the separate identity of each. Each will influence the other, extensively at times, but each is nonetheless distinct. Moreover, the person incumbent in a given principalship at a given time was, at some earlier time, a potential, rather than a practising principal, as shown by the designation on the left side of the box. The difference in status is associated with crossing the appointment threshold line, placed vertically toward the middle of the Figure. Movement across this line signifies a person's initial appointment to his or her first principalship. Once this happens, individuals will be in direct interaction with the task environment defined by the principalship(s) to which they are appointed, as indicated by the interactive arrow linking the principal and task environment boxes. The theory underlying the study holds that this direct interaction provides previously unavailable learning opportunities which will add to the store of task environment (domain relevant) schematic knowledge formed and held by the individuals concerned. This is shown by the descending arrow linking the interactive arrow to the polygon shaped box labelled as representing domain relevant schematic knowledge. The intervening box between the principal box and the domain knowledge model represents the complex of personal variables, such as intelligence, cognitive complexity, needs, and gender, which may influence how an individual learns from experience, and thus accumulates and integrates domain relevant knowledge.

Expertise model. The unshaded portion of this polygon to the right of the appointment threshold line represents the domain relevant schematic knowledge formed and held by a practising principal. Curved lines dividing the areas labelled High, Medium and Low represent variation in how well and rapidly individuals learn from their interactions with the task environment. A steeper learning curve results in more rapid and extensive integration of task relevant declarative and procedural knowledge associated with expertise; a flatter curve signifies less rapid and successful learning. Some task environments will presumably present greater and more apt

opportunities for learning than others, but while this is recognized here, it has not been incorporated into the model shown in Figure 1. Yet regardless of the learning opportunities available, the extent to which a principal (or any other role incumbent) benefits from direct learning opportunities will be mediated by his or her ability to learn, as illustrated by the bold arrow linking personal variables to the arrow linking interaction with the task environment to the expertise model. Leaving aside the learning opportunities associated with different task environments, then, the steepness of the learning curve governing the acquisition of expertise will be determined by personal variables such as cognitive ability, personal dispositions, aptitudes and the like. As shown by the arrow linking the personal variable box to the analysis area, we sought to collect data on several personal variables identified in the literature as being theoretically related to the development of administrative expertise. As described in greater detail later, these included several measures of cognitive complexity and the Ghiselli measures of the personality traits he identified as being associated with managerial talent.

The shaded, triangular portion of the expertise model extends to the left of the appointment threshold line in order to represent domain relevant knowledge acquired prior to an individual's first appointment to a principalship. As illustrated by the slope of this triangle, the amount of such knowledge can be expected to increase as potential principals prepare to try and pass the appointment threshold by engaging in both formal and informal learning activities. While formal instruction in school leadership and administration should obviously be expected to enrich an individual's specialized knowledge about this domain, it is clearly not the only source of relevant knowledge. Indeed, given the ubiquity of schools in contemporary society it seems reasonable to expect that people not preparing for the principalship will know something about the tasks and problems faced by principals. This may be particularly so for people who have had administrative experience of some kind in other kinds of organizations. As discussed at some length in the previous paper presenting the theoretical framework, our design called for the recruitment of a true novice group drawn from people who were not preparing for the principalship and had not otherwise been socialized to the work culture of schools. The "Novices" and "Pre-novice" boxes in the Figure represent the populations from which we recruited our true novice subjects The location of these boxes over the narrowing ends of the shaded triangle of the expertise model signify their likely possession of some, but probably not a great deal, of domain knowledge. The dashed lines descending from the novice boxes and from the principal box represent subjects recruited for the study. These dashed lines pass through the personal variables box,

acknowledging that the design called for the collection of measures of personal variables from all subjects. The lines also cut the expertise model at points indicative of the relative amounts of domain knowledge likely held by members of the novice and principal groups.

Data and analysis

The dashed lines descending from the subject groups all point to the circle labelled Sample Problem, signifying that all subjects were asked to think aloud about how to handle the problems presented in this case. As shown by the arrow connecting the task environment to the sample problem this was taken directly from an actual situation, modified so as to remove identifying information and condensed into brief case format. Participants' responses to the case were audio-taped and then transcribed. The concurrent think aloud [TA] transcripts thus generated constituted the main dataset for analysis, as shown in the Figure. As also shown by the appropriate arrows in the Figure, all transcripts were read and rated for the quality of the response to the case problem by a Peer Jury selected from experienced principals nominated by senior administrators of area school systems as described more fully in the appropriate section below. The jury ratings and evaluative comments made by jury members comprised additional data considered in the analysis.

While the design as summarized in Figure 1 provided for the comparative analysis of TA protocols from true novices and from practising principals, we also wanted to analyze responses from the principals according to the varying levels of expertise represented by the learning curves in the expertise model. The companion paper reviewing the theory base for the study discusses at some length the difficulties involved in attempting to identify more and less expert practitioners. Given the absence of readily available, reliable, comparative performance data for principals (and, indeed, all other administrators), use must necessarily be made of reputational data, preferably augmented by one or more independent indicators of proficiency. We nonetheless wanted to avoid or minimize the halo effects that would be generated by knowing who had been designated as more or less expert in advance of data collection. Our design solution was to gather reputational and performance data from co-workers after the TA and related data collection sessions were completed. This also allowed us to seek the informed consent of principals before collecting evaluative data, which facilitated the collection of a wider set of reputational data than might otherwise have been possible. These data were used in conjunction with the jury ratings to assign the participating

principals to high, medium and low imputed expertise categories, using the procedures explained more fully in a subsequent section of this paper.

Methods, procedures and descriptive data

Subjects

We set out to recruit 30 participants in each of the novice and principal categories. The principals were recruited by soliciting volunteers from selected school systems in south-western Ontario after obtaining permission from the Directors of Education (CEOs) of the systems concerned. Letters describing the study and inviting participation were sent to the principals of all elementary schools in the participating systems. The 31 who finally participated were selected primarily by matching their schedules to ours. Our data collection procedures for the principals required them to spend much of a single day with us, and as principals are busy people it proved impossible for some potential subjects to match their schedule with ours. The degree to which this may have biased subject selection is unknown, but as our only additional selection criteria were that the subjects were practising principals who volunteered, availability is unlikely to have systematically biased recruitment. The volunteer requirement will likely have had a more powerful and consistent effect, but this was unavoidable.

Recruiting subjects for the true novice group proved more difficult, far more so than we had anticipated. Volunteers were sought by advertising in the University newspaper and posting notices. Unfortunately, we had not budgeted for and thus did not offer to pay subject fees, and many of those responding to the advertisements declined to participate when informed of this. We eventually recruited 25 novice subjects, 10 of whom were B.Ed students and 17 graduate students or recent graduates from a variety of fields of study other than education, such as library science, arts and, in one case, biochemistry.

Demographic profiles of the participating principals and other subjects (the combined novice group will often be referred to as "others" in the papers that follow) are shown in Table 1. There are marked and statistically significant imbalances between the gender and age distributions of the two groups. Given the social history of the principalship and typical characteristics of university students, this was probably unavoidable. The gender imbalance, nonetheless, provides valuable opportunities to explore the effect of this variable in the data. As would be expected, there was also a significant imbalance in the administrative experience of principals and novices.

Table 1
Demographic profile of participants

	Novices		Principals		Total		Chi Squa
Gender	N	%	N	%	N	%	
Male	7	28	19	61	26	46	
Female	18	72	12	39	30	54	(1df) = 6.1
Totals	25	100	31	100	56	100	*p.* = .01
Age	N	%	N	%	N	%	
missing	1	—	0	0	1	—	
<= 30 yrs	15	63	0	0	15	27	
31-40	5	21	4	13	9	16	
41-50	4	16	21	68	25	46	[n=5
=> 51	0	0	6	20	6	11	(1df) = 32.3
Totals	24	100	31	100	55	100.00	*p.* < .00
Admin. Experience	N	%	N	%	N	%	
missing	3	—	0	0	3	—	
None	17	77	0	0	17	32	
<= 5 yrs	2	9	9	29	11	21	
6-10	2	9	13	42	15	28	
11-15	1	5	5	16	6	11	[n=5
=>16	0	0	4	13	4	8	(1df) = 31.9
Totals	22	100	31	100	53	100	*p.* < .00

Even so, five novices had had some administrative experience, three having gained six or more years of practical experience in the broader domain of administration.

Data collection

The principals spent much of a full working day with us on campus. The morning was devoted to completion of a battery of personal variable measures, the Symbol Card Task, and the think-aloud protocol session. After lunch, which we hosted as a token of our thanks, they participated in a structured interview probing their understanding of the principalship and involving the completion of a number of other tasks which are not reported in this symposium. Novice subjects participated in all of the activities which the principals completed in the morning session, but did so in a morning, mid-day or afternoon meeting as convenient to them. All data collection

took place in either the CASL Project office or an adjoining room in the John George Althouse Faculty of Education Building at the University of Western Ontario. The common activities in which all novices and principals participated are described in the following paragraphs.

Pencil & paper measures of Personal Variables. All participants were invited to complete the following scales and schedules:

1. A brief demographic questionnaire seeking information about age, gender and previous experiences.
2. Fiedler's Least Preferred Co-worker scale. A short, semantic differential instrument originally designed by Fielder (1958, 1967) to yield a measure of task or relationship oriented leadership style and thought by Bieri (1961) and others to provide a measure of cognitive complexity. Cronbach's alpha for the data obtained in this study was .905, with the overall mean score of 59.05 (s.d. 20.38), matching Posthuma's (1970) normative mean of 59.0. Additional information is provided in the accompanying paper reporting our findings with regard to cognitive complexity.
3. Ghiselli's Self Description Inventory [GSDI]. This is a disguised, 64 item forced choice inventory which is scored on 13 personality and motivational traits, nine of which appear to be related, to varying degrees, to what Ghiselli (1971) termed "managerial talent." A validation study by Schippmann and Prien (1985) supported continued use of the instrument as a "relatively quick and simple method of measuring occupationally related personality traits and abilities" (p. 1171). Cronbach's alpha for the eight scales Ghiselli considers to be most strongly related to managerial talent was .754, with the means and standard deviations on all 13 scales approximating his norms.
4. A 25 item Likert scale Leader Characteristics Questionnaire [LCQ]. We were particularly interested in incorporating a pencil-and-paper measure of Jaques' time-span of discretion construct as estimated by the Symbol Card Task and discussed in the description of this procedure given below. Sashkin's (1990) Leader Behavior Questionnaire [LBQ] includes a five item scale designed to measure this variable, but we did not want to use the entire 50 item LBQ. We consequently sought permission to use the 5 time-span items in a shorter instrument which also incorporated the self-efficacy and the power scales from the LBQ. The remaining 10 items were designed to investigate aspects of problem orientation. Data from this instrument are still being analyzed and will not be presented in this symposium, except for the scores from the time-span scales and responses to several global rating scales which were included at the end of the item schedule. The global rating scales are discussed further in the section dealing with the collection of reputational data.
5. Schroder's Paragraph Completion Test [PCI]. As discussed in Schroder and Suedfeld (1971) this is a relatively brief and open exercise believed to provide a measure of cognitive complexity. As explained in the accompanying paper on cognitive complexity, only 27 of the subjects in this study responded to this measure. Cronbach's alpha for the ratings on the 6 trails in this exercise was .836.
6. A forced choice values inventory, results from which are not considered in this symposium

Symbol Card Task. This is an activity developed by Stamp (1978) to provide insight into how an individual responds to a complex task. Jaques' (1976) original theory of time-span of discretion in work environments as subsequently refined and generalized in his (1982) Stratified Systems Theory [SST] holds that proficient performance in hierarchically superordinate administrative positions is related to an incumbent's ability

to effectively exercise discretion in the completion of tasks extending over increasingly longer time-spans. Under the theory, positions of responsibility requiring the direct direction and supervision of smaller work groups can be effectively managed by individuals functioning at Jaques' lower levels of work, which have time-spans of three months (Level I) and one year (Level II). The effective administration of larger and more complex organizational units is considered to require the administrators concerned to be able think and act within and over longer time frames, with qualitative breaks occurring at the two year, five year and ten year time-span levels. The symbol card task was developed to provide an opportunity of observing an individual's exercise of discretion in completing a complex, ill-defined task, and to thus estimate his or her time-span capability. The task was developed from Bruner's (1966; Bruner, Goodnow, & Austin, 1956) earlier studies of concept formation using symbol cards. As described more fully in the accompanying paper on cognitive complexity, subjects are asked to determine and then apply a rule for correctly sorting cards bearing circles, squares and triangles that vary in size, number and colour. Information about the rule they must seek is obtained by placing cards to match a standard display set and by processing limited feedback as to whether the placed cards do or do not match the rule. Data are obtained by observing how subjects approach and attack the task, and by counting the number of cards used and time taken to complete the task.

We incorporated the SCT into this study because of its theoretical relevance to the development of schematically organized domain knowledge, and particularly the steepness of the learning curves shown in the model of domain knowledge in Figure 1. As such, it can be viewed as another measure of cognitive complexity or, in Jaques' terminology, cognitive power. A related application concerns the levels of increasingly abstract thought that underlies the time-span of discretion theory and which, in turn, appears to be theoretically related to the richness and complexity of schemata formed and used by experts when working on domain problems. Some of our earlier work with the data from our first PPP study sought to relate a proxy estimate of level of abstraction to judged quality of response to the TA task (Allison & Allison, 1993). Encouraged by our findings, we sought to incorporate a more direct but domain free measure of this variable, which seemed best provided by the SCT. Results are reported in the paper on cognitive complexity prepared for this symposium.

Think-aloud session. The major dataset generated in the study was composed of transcripts of the think aloud sessions during which novice and principal participants responded to the sample problem presented in case format. Following a brief training session in the think-aloud process as recommended by Ericsson and Simon (1984), subjects were handed the case study and asked to read it aloud, interjecting their thoughts as they read, and then were encouraged to think out loud about how the principal in the case should proceed. When they finished, they were asked to recall all that they could remember about how they had thought about the case. They were then asked a series of standard debriefing questions about what they saw as the main problem(s) in the case, the goal(s) they were trying the achieve when searching for a solution, the main obstacles to be overcome in seeking to realize these goals, and how long they thought it would take to achieve them. All sessions were audio-taped and verbatim transcripts prepared. Machine malfunction resulted in the loss of one principal and one novice think-aloud session, leaving us with a dataset of 30 principal and 24 novice transcripts. Each transcript was edited to remove comments that might identify the subject, and each was assigned a random identification number. As described and discussed in two of the accompanying papers, the concurrent portions of the transcripts were segmented into individual thought units for analysis.

The case problem

The selection of domain problem(s) in novice-expert studies is an important design issue. When we began our first PPP study, we did not fully appreciate the theoretical importance of this consideration and simply decided to use the Miss Macdonald case that had been developed for and used in earlier studies by Nagy and P. Allison (1988). This case describes a situation in which Miss Macdonald, the long time teacher-librarian at Sugar Maple school, appears to be dissatisfied with her circumstances and is making only a lacklustre contribution to the school program. The immediately presented "surface" problem is Miss Macdonald's dissatisfaction and her apparent desire for a transfer to a classroom teaching position. Additional information points to less obvious, "deeper" concerns regarding the functioning of the library, staff relations, the school and its goals and program. Given the findings reported from previous expert-novice studies (e.g. Chi, Feltovich, & Glaser., 1981; Day & Lord, 1992), we anticipated subjects with greater expertise would notice the clues to potentially deeper aspects of the problem and pay more attention to these issues than would less expert subjects, who would tend to concentrate on the more

obvious surface aspects. To a degree, these expectations were substantiated. Yet when discussing results from the first study with colleagues we began to worry that the Miss Macdonald case would not be viewed as sufficiently ill-structured by more expert principals, and would thus not provide a good vehicle for isolating differences between how they and less expert principals responded to the case. After debating this concern for a while we concluded that substituting a case that more expert principals would likely see as more difficult would not necessarily yield more valuable data, especially as this also would handicap members of the novice group, and thus reduce the likelihood of collecting useful comparative data. What was required, we reasoned, was a case which was likely to viewed by most proficient principals as moderately rather than highly ill-structured. This would allow sufficient scope, we thought, for less proficient principals and at least some of our naive subjects to understand the problem and begin working on it, even though they would likely focus on the surface as opposed to the deeper features. And if, in consequence, more expert principals provided the kind of practised, automated response described as characteristic of the schema driven solutions to problems provided by experts in other domains, then so much the better. These considerations, together with the opportunity which would be provided to pool data from our first and second PPP encouraged us to retain the Miss Macdonald case in the present study.

We sought to test the degree of structure attributed to the case problem by using a brief problem sorting exercise following the think-aloud session. Subjects[1] were given eight cards, each of which contained a short, typewritten school-related problem. Two of these problems were adapted from those originally used in the Leithwood and Stager study (1989), others being drawn from a problem bank compiled during the course of the CASL Project. Following the procedure used by Leithwood and Stager, subjects were asked to read the problems to themselves and then sort them according how straightforward they thought the appropriate course of action would be, thus providing a estimation of the degree of structure they attributed to the problems. Once they had sorted the cards according to relative problem structure, they were then ask where they would place the Miss MacDonald problem relative to the others. Although specific rank placements varied, the principals never ranked the Miss MacDonald case as either the most or the least structured of the set, usually ranking it in or close to the middle of

[1] This problem-sort was added to our data collection activities part way through the study, and was occasionally omitted to save time. Consequently, we collected data from only 26 (47%) of the total 56 subjects, including 15 (48%) of the principals.

the array. Indeed, the mean rank accorded to the Miss MacDonald case by participating principals was 4.6 out of the total of 9 problems.

We did make some changes to the wording of case on the advice of the principals who sat on the jury judging think-aloud responses from the first study. These changes largely involved removing elements which the reputationally expert jurors identified as peripheral to the main problems. These changes had a more substantial effect on responses than we anticipated, limiting the extent to which we could pool transcript data from the first and second round studies. Because of this, the analyses reported in this symposium concentrate on data collected in the second round study, rather than pooled data from both as we had initially intended. An unanticipated positive consequence was that the use of two versions of the same case in the two data collection rounds provided an opportunity to examine how more peripheral elements in the version used in the first round acted as distractors. We hope to be able to report on these difference in the future.

Jury procedures

As shown in Figure 1, our design called for a jury of reputationally proficient principals to review, rate and comment on the responses to the case problem contained in the TA transcripts. Three jurors were selected from a pool of ten practising principals with whom we had worked on various training and professional development activities over the previous five years. Members of this pool were all originally nominated by the Directors of their school systems as potential instructors for a principals' certification course, all had completed graduate degrees, and all had, at the time of original nomination, served as principals for more than five years. The three jurors drawn from this pool were selected on the grounds of location and our own judgement. We wanted to ensure jurors were not sitting in judgement on colleagues with whom they worked, and thus members of the pool working in the same school systems from which we drew subjects were ineligible. Application of this exclusion criterion reduced the pool to six candidates, from whom we chose the three that we thought most suitable on the basis of having observed them at work in principals' training course and other activities. Two additional jurors were recruited on the recommendation of the Director of a school system outside of the region within which both subject principals and the other jurors were drawn. One of these jurors subsequently withdrew, and was replaced with a Doctoral student in educational administration who had served as an elementary principal for five years.

Each juror was provided with a binder containing the concurrent TA transcripts from all 54 subjects. identified by a randomly assigned identification number. Although the transcripts had been edited to remove names or other information that might have revealed the identity of the subject, it proved impossible to fully remove all comments revealing subjects' status without compromising the content of individual responses. Jurors were also provided with rating sheets on which they were asked to rate each transcript according to the overall quality of the response to the case on a scale from one to ten, where ten signified an excellent response. The jurors were also encouraged to provide brief written comments explaining their ratings in an appropriate area on the rating sheets. Finally, jurors were asked to declare whether they thought the person responding to the case was a non-principal, a beginning principal or an experienced principal by circling one of these options for each transcript.

Each juror privately reviewed and rated the cases individually over the course of eight weeks, and then mailed his or her rating sheets to us. Once all sheets had been received, summaries were prepared for each transcript showing the random identification number, the average of the five individual ratings, the standard deviation, range of ratings, and the rank and percentile statistics. Written comments were collated and typed on the summary sheets. All jurors then met together to review and agree on a consensus rating for each case. Each juror was provided with a set of the summary sheets and the transcripts, each of which was considered in turn. The discussion process was chaired by a member of the project team who facilitated discussion and ensured that divergent opinions were considered. Discussion was largely informal, but task oriented, focusing on strengths and weaknesses of each transcript. Jurors were able to reach consensus on appropriate ratings for each transcript. In doing so, 30 (72%) of the transcripts were awarded ratings lower than the mean calculated from the individual rating sheets, 13 (24%) transcripts received consensus ratings higher than the initial summary mean, and two (4%) transcripts retained the same rating.

Correlations between individual ratings from each juror ranged from .391 to .694: Cronbach's alpha for the correlation matrix was .876 Correlations between the final consensus ratings and the original individual ratings ranged from .635 to .875. Figure 2 plots the consensus jury ratings against their percentile rank. The plotted data have been jittered so that all cases are visible, and the plot is coded to show group membership. As would be expected, principals received the highest ratings, the upper quartile of the distribution being populated

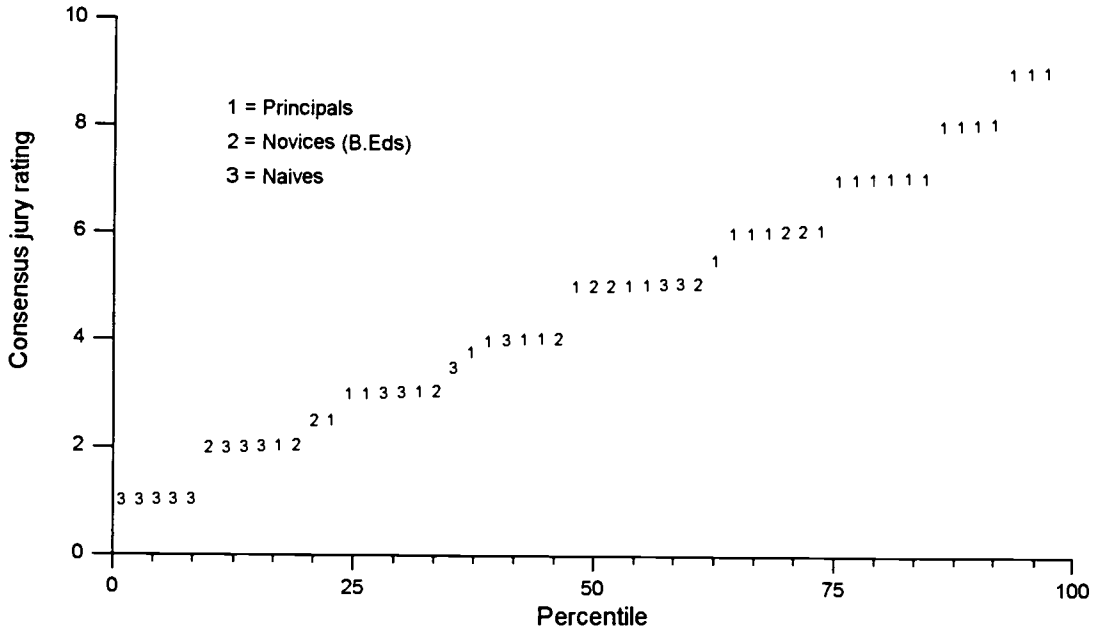

Figure 2

Plot of consensus jury ratings (jittered) showing subject group

exclusively by principals. But while, as would also be expected, the lower quartile contains mainly novices, three principals also appear here. Moreover, transcripts from three B.Ed. novices and one naive subject received ratings which placed them in the upper half of the distribution.

Imputed expertise categories

The jury ratings for the principal transcripts were used in conjunction with reputational and performance data collected from co-workers to assign each principal to one of three imputed expertise categories. This section first describes the reputational and performance measures used and then moves to explain how the imputed expertise categories were constructed.

Reputational data

The LCQ described previously included two scales which asked subjects to rate, on a ten-point scale, how well they thought they did their current job and their level of job expertise. At the end of the data collection

session, permission was sought from the principals to mail a version of the LCQ to six co-workers nominated by them, four teachers in their school and two school system administrators familiar with their work. All 31 principals agreed. An appropriately adapted version of the LCQ was thus mailed to 186 nominated co-workers with a cover letter explaining the procedure and a stamped, addressed return envelope to ensure anonymity. A total of 153 (83%) completed questionnaires were returned, representing an average of 3.3 responses from teachers and 1.7 from supervisory officers for each principal. In the version of the questionnaire completed by co-workers, respondents were asked to rate the principal named in the covering letter on the two global scales mentioned above, and to also rate the quality of the decisions made by the principal. Three co-worker ratings were thus obtained, one for how well the principal was seen as doing his or her current job, one for her or his level of comparative expertise,[2] and one rating the perceived quality of the decisions made by the subject.

Table 2
Descriptive statistics and correlations for self and co-worker ratings

		Job Performance		Comparative Expertise		Decision Quality		
	N	Mean	S.D	Mean	S.D	Mean	S.D	α
Self Ratings	31	7.90	1.09	7.83	1.42	—	—	.803
Teacher Ratings	101	8.51	0.78	8.47	0.93	8.46	0.82	.922
Superintendent Ratings	52	8.09	0.72	7.88	0.68	8.14	0.68	.931

Within Group Correlations for Mean Principal Ratings (N=31)

	Principal Self Ratings	Teacher Ratings		Superintendent Ratings	
	Expertise	Expertise	Decision Q	Expertise	Decision Q
Job Performance	.691	.803	.900	.682	.764
Comparative Expertise	—	—	.755	—	.726

All within group correlations significant beyond .001

Across Group Correlations for Mean Principal Ratings (N=31)

	Job Performance		Comparative Expertise		Decision Quality
	Self	Supers	Self	Supers	Superintendents
Teachers	-.007	.252	.278	.320	.191
Superintendents	-.066	—	.129	—	—

All of the across group correlations >.05 (non-significant)

[2] The actual wording of this item was as follows: "Considering all of the other elementary principals you have known, how would you rate the current level of expertise of the person named in the letter?", with the low end response scale being anchored with the phrase "Quite low: He/she is still something of a novice," and the high end with the phrase "He/she demonstrates a high level of expertise."

As summarized in Table 2, most of the ratings were high, with little variation, the largest standard deviations occurring for the self-ratings, reflecting the modesty of some principals. Cronbach's alpha for each of the respondent groups shows a high degree of internal reliability. One-way ANOVAs identified significant differences between respondent ratings for job performance [$F(2,181) = 3.86$, $p. = .023$] and comparative expertise [$F(2,181) = 3.88$, $p. = .024$] post-hoc analyses showing that in both cases the teacher ratings were significantly higher than the superintendent and the principal self-ratings. As would be expected from the high alpha co-efficients, correlations between the self and the mean co-worker ratings for each principal show high levels of agreement: principals who received high (or low) job performance ratings from their teachers or superintendents where quite likely to have also received high (or low) comparative expertise and decision quality ratings from same co-worker group. The strongest of these relationships—a notable .900—was between the mean teacher ratings for the job performance of their principal and their ratings for the quality of his or her decision making; the weakest was a still respectable .682 between superintendent ratings of principals' job performance and superintendent ratings of principals' comparative expertise. Principal self-ratings of their job performance and comparative expertise were reasonably strong ($r. = .691$), but while the principals saw these as being positively related, it seems clear that, as a group, they did not necessarily see one as highly dependent on the other.

Despite the generally high consistency within the respondent groups, the lower panel of Table 2 shows there was a substantial lack of agreement between the three groups as to the relative merits of the principals. This is most evident in the lack of agreement between the principals' self ratings of their job performance and those from both sets of co-workers: principals who rated themselves as doing a good job were not generally seen in the same light by either their teachers or their superintendents. There was higher agreement between the mean teacher and superintendent ratings of the principals' job performance, but the relationship remains weak ($r. = .252$) and fails to reach statistical significance ($p. = .172$). The highest across group correlation ($r. = .320$) shows a slightly stronger positive relationship for teacher and superintendent ratings for principal expertise, but this only approaches and does not achieve a respectable probability level ($p. = .079$). For our purposes in the current study, nonetheless, the most important message in Table 2 is the generally weak agreement between teacher and superintendent ratings on the measures of principal performance.

School effectiveness

In addition to the pencil-and-paper items completed by all subjects, the principals were also asked to complete the Index of Perceived Organizational Effectiveness [IPOE] as developed by Miskel (1982; Miskel, McDonald & Bloom, 1983) from Mott's (1972) earlier work. As shown in Hoy and Miskel (1991, Table 12.7) this is a brief eight item questionnaire intended to tap the eight dimensions identified by Mott as being central to organizational effectiveness. Two additional items were added in this study, one asking about the level of resources available in the school and the other asking for a global assessment of school quality, the specific wording of this item being as follows: "All things considered, how good is your school?" responses being made on a five point scale ranging from "Not very good" to "Very good." When we asked the principals for permission to send copies of the LCQ to co-workers we also asked if we could include the IPOE scale in the letter sent to their nominated teachers. All agreed. In this case the response rate was not as positive, with 89 (72%) of the nominated teachers returning usable questionnaires. Table 3 presents means, standard deviations and correlations for the IPOE (the sum the scores on the Mott scales) and the global "Good School" scale. While high (maximum feasible score is 40) the IPOE mean conforms closely to those reported by Miskel, McDonald and Bloom (1983) from their study of 89 public elementary and secondary schools in Kansas. The Good School global ratings are also high.

Table 3
Descriptive statistics and correlations for IPOE scores and Good School ratings

		IPOE			Good School item	
	N	Mean	S.D.	α	Mean	S.D.
Principals	31	30.92	3.40	0.76	4.58	0.49
Teachers	81	29.75	2.74	0.79	4.40	0.42
Pooled	114	30.31	2.48	0.80	4.52	0.35

Within Group Correlations (N=31)

	IPOE	
	Principals	Teachers
Good School	.785	.650

p. <.001 in both cases.

Across Group Correlations (N=31)

	Principals	
Teachers	IPOE	Good School
IPOE	.358	.153
Good School	.313	.032

p. > .05 in all cases.

with tight standard deviations, signifying that respondents generally rated their schools as being very good, principals tending to rate their schools slightly higher on both the IPOE scales and the Good School item. The correlation of .649 ($p.$ = .0002) for the mean teacher ratings for each school indicates reasonable but by no means substantial agreement between their responses to the IPOE and the Good School item, the comparable correlation coefficient for the principals indicating slightly higher agreement. Once again, however, there was markedly lower agreement between the mean teacher ratings for each school and the principal's ratings, $r.$ values ranging from .032 to .358, none reaching significance.

The imputed expertise categories

Ericsson and Charness (1994, p. 732) caution that "researchers cannot seek out experts and simply assume that their performance on relevant tasks is superior; they must instead demonstrate this superior performance." This appears particularly pertinent when the search for experts must necessarily rely to some extent on reputation, as will be the case in most social action domains such as administration. With this in mind, our design provided for the use of both reputational data and peer jury ratings of how well subjects had responded to the case study task. We had originally intended to construct a single omnibus measure from the co-worker ratings to serve as a reputational indicator which would be used in conjunction with the jury ratings to identify principals who were highly regarded by their co-workers and were judged to have provided a high quality response to the case problem. We regarded the IPOE scores as a potentially less valuable indicator given the influence of environment on school effectiveness, but thought that it, together with the Good School ratings, should provide potentially useful additional information. As it turned out, and as shown by the correlations in Table 2, the ratings from the teachers and superintendents could not be sensibly pooled, but had to be considered separately. So too with the IPOE and Good School score, where it also seemed prudent to consider only the teacher ratings.

We proceeded by summing the mean job performance, comparative expertise, and decision quality ratings from the teachers and then the superintendents. The resulting scores for these two co-worker groups were then ranked and transformed into percentiles. Percentile scores were also calculated for IPOE scores and the Good School ratings from the teachers, and the consensus jury ratings. The five sets of percentiles were then tabulated and inspected. None of the principals had percentile scores at or above 75 on all five distributions, although six

scored above the 75th percentile on three out of the five scales, and seven scored above this mark on two scales. No less than eight principals fell below the 75th percentile on all of the five comparisons. Yet while it was evident that some principals outshone others, there was no simple pattern that could be used to mechanically select a more highly rated, arguably expert, group.

We eventually adopted a set of decision rules which allowed different weights to be assigned to the various ratings. In order to be included in the High imputed expertise group, principals had to satisfy three contingent criteria. First, their transcripts had to have received a rating of seven or above from the peer jury. This we justify on the ground of satisfying Ericsson and Charness' requirement for demonstrated expertise. Thirteen of the 30 principals (one transcript had been lost due to a machine malfunction) satisfied this criterion. Second, they had to have received relatively high ratings (=> 75th percentile) from *either or both* of the co-worker groups (teachers *or* superintendents}. Eight of the thirteen principals with jury ratings of seven and above met this criterion. Third, one or both of the school effectiveness and quality ratings had to be at or above the 75th percentile, *unless* a principal had received high ratings from *both* teachers and superintendents. The condition attached to the third criterion was adopted to assign greater weight to the co-worker ratings than the school effectiveness and quality ratings: a principal rated above the 75th percentile by both his or her teachers and her or his superintendents would, we reasoned, have a good claim to expertise even if his or her school was not rated in the upper quartile on the school measures, which might be attributable to environmental or other contingent circumstances. Five of the eight principals satisfying the first two criteria had either school effectiveness or quality ratings above the 75th percentile, none having sufficiently high scores on both. One additional principal was judged to qualify for the final group on the grounds of having high ratings from both teachers and superintendents. The IPOE score for this principal's school fell at the 30th percentile and the Good School rating at the 48th percentile. In sum, six principals passed through the three screens devised to select members of the high imputed expertise category. Expressed another way, slightly more than half of those that satisfied the first criterion of a transcript rating of seven or higher failed to pass the other two screens.

We must admit to not being entirely satisfied with the procedures described above, and intend to explore the ratings data further in search of a cleaner and preferably simpler way of differentiating between higher and lower rated subjects. It may be sensible, for instance, to give greater weight to the co-worker ratings of comparative

expertise, rather than combining these ratings with the job performance and decision quality ratings. Even so, we feel the procedure adopted above is arguably valid and robust. Multiple data sources were used and an attempt was made to incorporate and balance estimates of performance on a specific task together with ratings of more general job performance. The difficulty, as always, is where to draw the lines. In this case we are not certain this one line has been drawn correctly, but we are as sure as we can be from the data available that the members of the High impued expertise group have good claims to be recognized as more proficient than more than a few of the other principals in the study, especially those included in the group described next.

Scanning the percentile data it was evident that a few principals had received consistently low ratings. One of those receiving a low jury rating, for example, had IPOE, school quality, and co-worker scores which all fell below the 20th percentile, and several other principals approached this nadir. When we began the study, we were not committed to isolating a group of relatively weak principals, being quite prepared to adopt the convention of recognizing more and less expert principals used by the Leithwood group. Given the pattern evident in the percentile ratings data, however, we could not justify including those with consistently poor ratings within a residual Medium imputed expertise group. Two criteria were applied to formally identify members of the Low expertise group. The first approximates the three strikes and out rule: anyone falling below the 25th percentile on three of the four external ratings (teacher, superintendents, IPOE and school quality) was classified in the low imputed expertise group. To honour the principle of assigning greater weight to the co-worker ratings of principals used when selecting members of the High expertise group, the second criterion specified that anyone with *both* teacher and superintendent ratings below the 25th percentile was also placed in the low group. The jury ratings were not considered when applying these rules. We reasoned that while a high jury rating logically implies the possession of expert knowledge, a low rating does not necessarily indicate a lack of expertise, perhaps just a failure to demonstrate it within the limitations imposed by the think aloud task. Five principals were classified in the Low imputed expertise group, leaving 19 principals in the residual Medium category.

Conclusion

This paper has described the conceptual design, data collection methods and other procedures used in the study of principal problem processing reported in the papers prepared for this symposium. Three analytical papers follow, two of which present analyses of the transcript data and one some results from the various measures of cognitive complexity incorporated in the design. A final paper draws some implications for the preparation and professional development of school principals.

References

Allison, D. J., & Nagy, P. (1991). A study of principal problem solving: An introduction the study. Paper presented to the Annual Conference of the American Educational Research Association. Chicago. (ERIC Document #ED 333 581).

Allison, D. J., & Allison, P. A. (1993). Both ends of a telescope: Experience and expertise in administrative problem solving. *Educational Administration Quarterly, 29*, 302—321.

Bieri, J. (1961). Complexity—simplicity as a personality variable in cognitive and preferential behavior. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience*. Homewood IL: Dorsey Press.

Bruner. J .S. (1966). *Toward a theory of instruction.* Cambridge, MA: Belknap Press.

Bruner. J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: John Wiley.

Chi, M.T.H., Feltovich, P. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Cowan, D. A. (1990). Developing a classification structure of organizational problems: An empirical investigation. *Academy of Management Journal, 33*, 366-390.

Day, D. V. & Lord. R. G. (1992). Expertise and problem categorization: The role of expert processing in organizational sense-making. *Journal of Management Studies, 29(1)*, 35—48.

Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition, *American Psychologist, 49*, 725—747.

Fiedler, F. E. (1967). *A theory of leadership effectiveness.* New York: McGraw Hill.

Fiedler, F. E. (1958). Leader attitudes and group effectiveness. Urbana, IL: University of Illinois Press.

Ghiselli, E. E. (1971). *Explorations in managerial talent.* Pacific Palisades, CA: Goodyear Publishing Company.

Hallinger, P., Leithwood, K., & Murphy, J. (Eds.). (1993). *Cognitive perspectives on educational leadership.* New York: Teachers College Press.

Hoy, W. K. & Miskel, C. G. (1991). *Educational administration: Theory, research and practice* (4th ed.). New York: McGraw Hill.

Jaques, E. (1976). *A general theory of bureaucracy.* London: Heinemann.

Jaques, E. (1986). The development of intellectual capacity: A discussion of stratified systems theory. *The Journal of Applied Behavioral Science, 22*(4), 361—383.

Leithwood, K. A., & Stager, M. (1989). Expertise in principals' problem solving. *Educational Administration Quarterly, 25*, 126—161.

Leithwood, K. & Steinbach, R. (1995). *Expert problem solving: Evidence from school and district leaders.* Albany, NY: State University of New York Press.

Miskel, C. G. (1982). *Structural linkages, expectancy climate and school effectiveness.* Research Report No. 58. Institute for Research in Learning Disabilities, University of Kansas at Lawrence

Miskel, C. G., McDonald, D., & Bloom, S. (1983). Structural and expectancy linkages within schools and organizational effectiveness. *Administrative Science Quarterly, 19*(1), 49—82.

Mott, P. E. (1972). *The characteristics of effective organizations.* New York: Harper and Row.

Nagy, P. & Allison, P. A (1988). School-level decision making: A cognitive perspective. Paper presented at the annual meeting of the Canadian Society for the Study of Education, Windsor, Ontario.

Posthuma, A. B. (1970). Normative data on the Least Preferred Co-worker scale and group atmosphere. University of Washington, Seattle. Department of Psychology, Organizational Research Technical Report 70-B.

Reimann, P. & Chi, M. T. H. (1989). Human expertise. In K. J. Gilhooly (Ed.), *Human and machine problem solving,* (pp. 161—191). New York: Plenum Press.

Sashkin, M. (1990). *Leader Behavior Questionnaire*. (Third Ed.). King of Prussia, PA: Organizational Design and Development.

Schippmann, J. S., & Prien, E. P. (1985). The Ghiselli self-description inventory: A psychometric appraisal. *Psychological Reports, 57*, 1171—1177.

Schroder, H. M. & Suedfeld, P. (1971) *Personality theory and information processing*. New York: Ronald Press Co.

Srivastva, S. (1983). *The executive mind*. San Francisco: Jossey Bass.

Stamp, G. (1978). Assessment of individual capacity. In E. Jaques, R. O. Gibson, D. J. Issac (Eds.) *Levels of abstraction in logic and human action: A theory of discontinuity in the structure of mathematical logic, psychological behaviour and social organization* (pp. 251—270). London: Heinemann.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science*, (pp. 527-579). Cambridge, MA: MIT Press.

List of Papers Comprising the 1996 AERA CASL Symposium Set

Author(s)	Titles
Allison, D. J.	Problem processing and the principalship: Theoretical foundations and the expertise issue.
Allison, D. J. & Morfitt, G.	Problem processing and the principalship: Design, methods and procedures.
Morfitt, G., Demaerschalk, D., & Allison, D. J.	Paying attention: Content considered by experts and others when solving a case problem.
Allison, P. A., Demaerschalk, D., & Allison, D. J.	Thinking through an administrative problem: Differences between expert, average and true novice responses.
Allison, D. J., Morfitt, G., & Demaerschalk, D.	Cognitive complexity and expertise: Relationships between measures of cognitive complexity and abstraction and responses to a case problem.
Allison, P. A., & Allison, D. J.	Cognitive approaches to school leadership: Some implications for the selection, training and development of principals.

Contact address: The CASL Project
Faculty of Education
The University of Western Ontario
London, Ontario, N6G 1G7
Canada.

email: allison@edu.uwo.ca

EA028656

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Problem processing and the principalship: Design, methods and procedures	
Author(s): Derek J. Allison and Grace Morfitt	
Corporate Source: University of Western Ontario	Publication Date: April 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

	The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2 documents	
[X] Check here For Level 1 Release: Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy.	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	[ ] Check here For Level 2 Release: Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.
	Level 1	Level 2	

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please	Signature: Grace Morfitt for	Printed Name/Position/Title: Derek J. Allison Associate Professor	
	Organization/Address: University of Western Ontario Faculty of Education 1137 Western Rd. LONDON, Ont. N6G 1G7	Telephone: (519) 661-3182	FAX: (519) 661-3833
		E-Mail Address: allison@edu.uwo.ca	Date: Oct. 21/96 July 30/97