

# A Study of Administrative Expertise in Participant Performance on the NASSP Assessment Center

MILES T. BRYANT

*Department of Educational Administration, University of Nebraska-Lincoln, 1204 Seaton Hall,  
Lincoln, NE 68588-0638*

## Background and conceptual framework

School districts are using increasingly the National Association of Secondary School Principals (NASSP) Assessment Center as a predictor of future job performance.<sup>1</sup> This assessment process requires a prospective administrator to perform a number of simulations while being observed by trained assessors to obtain a score that can be used to determine the present or predicted level of administrative skill expertise. South Carolina, Missouri, and California are among states where such centers are in use as a selection screen.<sup>2</sup>

Career advancement in public education is predicated normally upon an accumulation of knowledge (college course work) and experience (years of service). This acquisition of experience and education forms the major portion of the evidence used to characterize an individual as one who can be assumed to possess administrative expertise. That assumed expertise, real or otherwise, becomes the basis for career advancement as an administrator. The logical conclusion by local hiring authorities is that experienced administrators are likely to perform better than inexperienced ones.

One element of expertise is domain-relevant knowledge (Amabile, 1983).<sup>3</sup> No matter what the field of endeavor, it is unlikely that one will achieve levels of expertise in the absence of knowledge about that field or domain. Thus, the first requirement in gaining expertise is learning or acquiring domain-relevant knowledge. Lesgold (1983, p. 3) argued that in the acquisition of domain-relevant knowledge there are two types of learning—"knowledge which comes from variation and knowledge which comes from repetition." Lesgold used the chess player to illustrate how these types of knowledge are acquired, but his concepts should hold across other areas of human activity. Thus, the educational administrator can be expected to acquire knowledge both from facing a variety of unique situations and from repetition of organizational routines that are incrementally modified.

To the degree that participants in the NASSP Assessment Center have experienced greater administrative variety and repetition, they should do better in their performance of certain skill dimensions assessed in the simulation exercises. If the basic premise that the Assessment Center measures administrative skills is accepted, greater experience should result, *ceteris paribus*, in higher performance ratings.<sup>4</sup>

The NASSP Assessment Center captures and measures performance on 12 administrative behaviors: problem analysis, judgment, organizational ability, decisiveness, leadership, sensitivity, stress tolerance, oral communication, written communication, range of interest, personal motivation, and educational values (Wendell, 1988). The first six of these dimensions are considered administrative skills; these dimensions are of particular concern to the investigator.

Because the center process requires that participants enact given roles in a simulated administrative situation while being observed, it can be hypothesized that participants who have some experience will perform better on simulations that require administrative skills (the first six dimensions). Conversely, there is no particular reason why the performance of administratively experienced and inexperienced participants should differ on interpersonal, communication, or personal dimensions (the latter six dimensions). The acquisition of expertise in these latter areas deals more with real-life experiences than the administration of educational organizations.

There is a substantial literature comparing experience (experts) to inexperience (novices), a literature that leads to the common sense notion that experienced performers do better than inexperienced ones (Bryan & Harter, 1897; de Groot, 1966; Chase & Simon, 1973; Newall, 1973; Wiedenbeck, 1985). A central theme in this literature is that experts can visualize the temporary patterns of the present. Experts possess a domain-relevant knowledge that allows them to recall more detail than the less expert (Voss, Vesonder, & Spilich, 1980). Consequently, they are able to identify the essence of the problem situation they face. Chase and Simon (1973) refer to this practice as chunking.

Voss and associates (1980) maintained that experts are able to utilize domain-relevant knowledge to process and control information more accurately. Voss and associates (1980) studied baseball fans, asking experienced and inexperienced groups to produce a narrative account of a half inning of baseball. Members of the low knowledge group were less able to recall correctly the sequence of events that had occurred.

In another study aimed directly at problem solving, Voss and colleagues (1984) analyzed the reasoning characteristics used by political scientists to deal with a hypothetical problem concerning Soviet agriculture. The reasoning of these political scientists was contrasted with that of two other groups: chemists and undergraduate political science majors. Voss and colleagues found that the expert political scientists made fewer errors and provided detailed support for their recommendations while the less expert made more errors in judgment and did not provide the same degree of support. Lesgold (1983) refers to this same problem-solving process with the terms *representation* and *solution*. Experts are better able to represent the problem, i.e., to frame it and then to construct a solution.

In a study of the dynamics of reading, Kintsch and van Kijk (1978) found that persons with high domain-relevant knowledge had a better short-term memory and were able to process more main-point information. Wiedenbeck (1985) argued that expert computer programmers were able to isolate better the core of a complicated computer program whereas nonexperts stumbled over the control language that typically surrounds the purposeful part of the total program. Wiedenbeck's finding is consistent with one of the central themes of this literature: in the problem arena, experts are more accurate and efficient in defining the key issues.

What is true for chess players, computer programmers, baseball fans, and others should be true for educators. Those who have acquired some degree of expertise should be more adept at solving educational problems than those who have less expertise. Berliner (1987) reviewed work that has attempted to contrast expert and inexpert teaching performances. He reported work by Hanninen that found the following two characteristics: (1) expert teachers used a "higher-order system of categorization in analyzing problems"; and (2) expert teachers have a "richer problem representation" than novice teachers, i.e., they find more meaning in incidents and are better able to articulate that meaning.

Berliner (1987) also reported work in progress on expert and novice teachers as they respond to a micro-teaching situation in an unfamiliar setting. Here expert teachers reported feelings of discomfort because they were unable to prepare and control the content of the lesson and because, lacking knowledge of these new students, they felt unable to predict probable student behavior. Novice teachers did not report being troubled by the setting or lack of knowledge. This finding echoes the conclusion from other fields that one characteristic of expertise is the ability to anticipate events based on a knowledge of patterns.

### **Purpose of the study**

Using the literature on expertise for a theoretical grounding, the purpose of this study was to see if NASSP Assessment Center participants with administrative experience achieved higher ratings than those with little or no administrative experience. It was assumed that concepts discussed previously would hold true in the field of educational administration. Thus, those with experience should display characteristics of expertise similar to those of actors in other areas of knowledge:

1. Experienced administrators should process the information given them in the simulation more quickly, more accurately, and more efficiently (Voss et al., 1980, 1984; Weidenbeck, 1985).
2. Experienced administrators should be able to represent problems more completely and totally (Lesgold, 1983).
3. Experienced administrators should display a better short-term memory and better identify main points of simulated problems (Kintsch & Van Kijk, 1978).

Consequently, the scores or ratings of experienced administrators should provide evidence of greater processing ability, more accurate representation abilities, and more successful identification of various simulation problems.

The actual process followed in an Assessment Center allows one to test these hypotheses. For example, participants are required to play the role of an administrator responding to a series of items in an in-basket. Each item has particular responses that are tied to skill dimensions. If, for example, the in-basket contains a memo requesting that a well-placed civic group would like to use the school's gym, one of the responses is to check board policy, which in turn is tied to the skill dimensions of problem analysis or judgment. The participants that indicate in writing that they would check school board policy will receive a higher rating than those who do not. The rating that participants receive is the result of multiple assessments by trained evaluators and forms the dependent variable used by the investigator to capture the impact of experience.<sup>5</sup>

### **Sample and methodology**

In order to compare the performance of experienced and inexperienced educators, data were gathered from participants who had been assessed at the NASSP Assessment Center at the University of Nebraska–Lincoln. Eighty educators—the total number of participants at the point when the study commenced—formed the study group. Of these 80, 28 indicated primary experience at elementary level, 31 indicated primary experience at the secondary level, and 21 indicated experience in higher education or some other level of education. In terms of administrative experience, 28 percent reported none, 35 percent indicated five years or less, and 37 percent indicated over five years of experience.<sup>6</sup>

The design used to compare experienced with inexperienced administrators falls into what Campbell and Stanley (1963) refer to as the static-group comparison. In such a design, one group that has experienced the item of interest to the investigator is compared to another group that has not experienced the item. In this study that item was administrative experience. This study thus possesses the flaw of selection bias characteristic of such an approach. If differences between experienced and inexperienced participants were found, there was no way of knowing if there might not be some unusual characteristic at work in the self-selection process by which participants come to the NASSP Assessment Center. Indeed, if one considers the universe from which participants in the Assessment Center process were self-selected, the fact that subjects in this study self-selected themselves presents problems for internal validity. The implication of this design flaw for the findings of the study is discussed in the concluding remarks.

The study was organized to compare the performance of participants at different levels of administrative experience using the behavioral skill dimensions of the NASSP Assessment Center. T-tests and analysis of variance were used to determine if a significant difference existed between groups with varying degrees of experience.

*Table 1.* Means and standard deviations for administrative skill dimensions by experience.

Skill Dimension	Group 6 (0–5 Years)			Group 5 (6–15 Years)		
	N	Mean	StD	N	Mean	StD
Problem Analysis (PA)	45	27.7	5.75	30	31.2	4.58
Judgment (J)	45	27.0	7.17	30	28.0	6.22
Organizational Ability (OA)	45	28.1	8.01	30	31.3	6.01
Decisiveness (D)	45	33.1	8.43	30	33.6	8.34
Leadership (L)	45	30.7	8.80	30	32.5	6.80
Sensitivity (S)	45	30.3	7.5	30	33.4	6.00

### Analysis

Participants were separated into comparison groups with differential levels of administrative experience as follows:

- 1) Group 1 = 2–15 years  
Group 2 = 0–1 years
- 2) Group 3 = 4–15 years  
Group 4 = 0–3 years
- 3) Group 5 = 6–15 years  
Group 6 = 0–5 years
- 4) Group 7 = 8–15 years  
Group 8 = 0–7 years

Generally, there were slightly higher mean scores for those participants with more years of administrative experience. For example, table 1 presents descriptive statistics of the scores achieved for two groups of participants on six administrative skill dimensions. As may be expected, participants with more than five years of administrative experience achieved slightly higher mean ratings on all of the administrative skill dimensions.

Having established groups and determined the mean ratings for participants in different groups on the six dimensions, the next step was to test if the differences in the mean values between groups were statistically significant. T-tests were used.

The higher rating on the problem analysis score of participants in Group 5 (6–15 years of administrative experience) was found to be statistically different ( $p < .05$ ) than that of participants in Group 6 (0–5 years of administrative experience). This finding was repeated in the analysis of the difference in rating between Group 7 (7–15 years of administrative experience) and Group 8 (0–6 years of experience).

Significantly, this result was obtained only with the one dimension presented in

table 2—problem analysis. Differences between the experienced and inexperienced participant on the other administrative skill dimensions—judgment, organizational ability, decisiveness, leadership, and sensitivity—were insignificant. There were also no significant differences in the nonadministrative skill dimensions.

To test the effect of experience on participant performance further, a one-way analysis of variance was conducted. Again, groupings were made according to experience. This analysis found that ratings achieved by experienced participants (five or more years of administrative experience) varied significantly from those of inexperienced participants (five or less years of administrative experience) on two skill dimensions: problem analysis ( $F=4.90$ ,  $df=1,78$ ,  $p<.03$ ) and organizational ability ( $F=4.92$ ,  $df=1,78$ ,  $p<.03$ ). An analysis of the variance in the other administrative, interpersonal, and communication skill dimensions again yielded no significant values.

*Table 2.* Differences in participant performance on problem analysis by years of administrative experience.

Groups	N	Mean	SD	t
Goup 1 (2–15)	58	29.66	5.60	1.66
Group 2 (0–1)	22	27.36	5.25	
Group 3 (4–15)	56	28.93	5.27	-.20
Group 4 (0–3)	24	29.21	6.41	
Group 5 (6–15)	36	30.67	4.85	2.51*
Group 6 (0–5)	44	27.68	5.81	
Group 7 (8–15)	28	30.82	4.05	2.43*
Group 8 (0–7)	52	28.06	6.05	

## Discussion of findings

The NASSP Assessment Center process holds the promise of becoming a widespread mechanism used by public school districts to determine one's fitness for future administrative employment.<sup>7</sup> Participants rated highly by assessors will be viewed as possessing the administrative skills that make them desirable candidates. Those rated as poor administrative prospects will be judged inexpert or unlikely candidates. Policy issues accompany such a development in personnel selection approaches. How good is the screen? What is lost because the center only simulates the educational setting? Have all of the elements that combine to produce expertise been identified? Do all participants have an equal opportunity to excel or are there particular biases that favor some?<sup>8</sup>

If the NASSP Assessment Center is to be trusted as an employment screen, it should be able to capture the accumulation of expertise that, as was argued earlier, is the logical byproduct of the knowledge gained through varied and repetitive experience.

The results of the investigator's analysis above would suggest the NASSP Assessment Center does not capture the assumed expertise (experience) possessed by participants on many dimensions.<sup>9</sup> The Assessment Center process does appear to distinguish between experience and inexperience on one, probably two, dimensions: problem analysis and organizational ability. There were no statistically significant differences in the ratings of participants with administrative experience and those without administrative experience on any of the other skill dimensions.<sup>10</sup>

### **Policy implications**

If the study reported above were alone in raising questions about the predictive accuracy of the NASSP Assessment Center, the growth in its use as an employment screen might be viewed with some equanimity. A study of 80 participants in one state at one center cannot be generalized. A review of one other attempt to understand the validity of using the assessment process as a means of identifying administrative expertise (or talent), however, also raises questions about its use as an employment screen.

In 1979 the NASSP commissioned a three-year validation study of the Assessment Center process. Schmitt and associates (1982) of Michigan State University conducted a study that attempted to relate job performance evaluations to Assessment Center ratings. Assessment Center participants subsequently hired for administrative positions were rated on 15 behaviorally anchored rating scales.<sup>11</sup> Rating scales were obtained from the participant's self-evaluation, from supervisors, from teachers, and from support staff. These ratings then were correlated with the earlier scores the same administrator had achieved as an Assessment Center participant. The design was to see how well the Assessment Center predicted on-the-job performance. A summary of the correlations obtained in this study has been compiled in table 3.

Table 3 tallies the correlations obtained by Schmitt and associates of the relationship between job performance evaluations and Assessment Center ratings. Several interesting aspects emerge when these correlations are arrayed according to strength.

First, neither the self-evaluations nor the evaluations of support staff had a strong relationship with Assessment Center scores. For the self-evaluations, 86 percent (66 + 101/195) of the items were correlated at .15 or less. Eighty-two percent (71 + 88/195) of the items correlated between support staff ratings and Assessment Center ratings were at or below .15. The percentage of stronger correlations (.16 to .45) increased when superior ratings and teacher ratings were correlated with Assessment Center ratings. Still, the bulk of the correlations were .15 or less (59 percent for supervisors and 66 percent for teachers). What may be observed here is a Rashomon effect where one's version of reality differs depending upon how one was involved in the crime.<sup>12</sup> Clearly, agreement across groups about administrator performance varies, a fact that raises questions about the diagnostic accuracy of the assessment process.

Table 3. Summary table of validity study correlations between assessment center ratings and ratings by evaluation groupings.

Correlation Range	-.50 to -.1	.0 to .15	.16 to .30	.31 to .45	N
When Self-Rated					
Item Totals	66	101	28	0	195
Significant Items*	15	0	26	0	41
When Supervisors-Rated					
Item Totals	36	79	68	12	195
Significant Items*	1	1	59	12	73
When Teachers-Rated					
Total Items	38	91	59	7	195
Significant Items*	4	0	40	6	50
When Support Staff-Rated					
Total Items	71	88	33	3	195
Significant Items*	5	0	12	3	20

\* Statistically significant at  $< .05$ .

Source: Schmitt et al., 1982 pp. 28-31.

A second interesting aspect of the tallies in table 3 is that although correlation items in the upper ranges tended to be statistically significant ( $p < .05$ ) more frequently, in none of the rated groups was a majority of correlations significant. When supervisors rated participants, 37 percent (73/195) of the items were significant. When evaluated by self, by teachers, and by support staff, the significant items became a lesser percentage of the total. Even though on certain items there was a moderate (in the .16 to .45 range) relationship, the conclusions one must reach about the total diagnostic accuracy of the Assessment Center need to be modest in light of these data. A competent administrator in an hour-long interview might predict with just as much accuracy the likelihood of a candidate's performing effectively as a future school principal.

A third interesting aspect of these data is that the relationship between what workplace colleagues in a school said about an administrator's ability and what the Assessment Center predicted they would say about that administrator's ability is not particularly strong. Few correlations were over .30. It is not surprising that ratings by superiors have the strongest relationships; superiors possess the bias of initially having selected that administrator. It is not surprising that the ratings of support staff (janitors, secretaries, etc.) have a weak relationship; typically such personnel have the least opportunity to observe the complexity of administrative work and thus will have the least information upon which to base judgment. Given a growing conviction among practitioners that the Assessment Center process is a good predictor of future administrative success, it is surprising that the relationships between the on-the-job evaluations (after the fact) and the Assessment Center evaluations (before the fact) are not stronger.



## Conclusions and recommendations

What is similar about the findings of the NASSP-commissioned study and this comparative investigation of experienced and inexperienced performance on the Assessment Center is that neither study can lead to sanguine conclusions about the capacity of the Assessment Center process to predict administrative performance. It apparently will predict well on a few dimensions (problem solving and organizational ability).

This questionable capacity to predict future performance does not deflate the worth of the Assessment Center process. What is implied is that it is better used for purposes other than selection. For example, the Assessment Center provides a participant with a great deal of material about how others view his or her performance in a simulated situation. This information can be used as an heuristic device to generate questions and directions about an individual's work and self-development. In connection with graduate study, the results of an assessment center could be used to help a graduate student in administration select particular course work.

It was pointed out earlier that the research design of the investigator's study is flawed. However, an argument can be made that in a study where practicing administrators have self-selected themselves as subjects, the logical prediction would be that the effect of expertise should produce more difference between experienced and inexperienced. Administrators lacking confidence in their administrative abilities would not have been likely volunteer participants in the Assessment Center. Therefore, one can argue that those who volunteered should be a more able group than one would obtain in a random selection process. More rigorous research on the relation of experience to performance is needed. Research is needed to study whether the Assessment Center's significant responses promote a narrow vision of administrative success. More research is needed about the wisdom of policy decisions that utilize Assessment Centers as an administrative employment screen.

In the last analysis, what William Martin, formerly of the Federal Reserve Board, said about statistics may capture what we should believe about the Assessment Center process: It "should be used like the drunk uses a lamppost—for support, not illumination."

## Notes

1. For a thorough description of the history and nature of the NASSP Assessment Center see Wendel, (1988).

2. Missouri, for example, now requires all prospective administrators to "pass a special Assessment Center before being eligible for their professional certificates" (see *NASSP Newsletter*, May 1987, p. 4)

3. According to Amabile, domain relevant skills include factual knowledge, technical skills, and special talents (Amabile, 1983, p. 67).

4. After a certain point, scores on the assessment center will begin to be less related to experience as a regression factor begins.

5. The NASSP Assessment Handbook which is used by assessors provides complete information about the assessment process. It is prepared by NASSP, 1904 Association Drive, Reston, Virginia 22091.

6. In the early development of the assessment center at the University of Nebraska-Lincoln, practicing school administrators were encouraged to participate as a means of introducing the technology to the practitioner community. It has become increasingly common to disallow the participation of practicing administrators.

7. NASSP officials estimate that about 7,000 individuals have participated in the assessment process and that another 3,000 educators are trained as assessors.

8. Mary Ford (1987) found that women attach more to the Assessment Center process than do men. Schmitt and associates found that white men and women outperform minority men and women.

9. The author reminds the reader at this point of the earlier argument equating expertise with experience, holding other variables such as intelligence and capacity equal.

10. This finding may not agree with conclusions reached by Schmitt and associates who wrote: "There is some tendency for non-teaching personnel to perform better than other groups" (Schmitt et al., 1982, p. 18). To the contrary, it appears that one only can support this claim on two dimensions—problem analysis and organizational ability. If this is what is meant by the phrase "some tendency," then there is agreement. And if this is not what is meant, the conclusions of this validation study are misleading, for it labels the Assessment Center process as a valid selection device.

11. Schmitt and associates, 1982, 20.

12. *Rashomon* is a Japanese classic by Ryunosuke Akutagawa in which a murder is recounted from three perspectives: that of an outlaw named Tajomaru, that of the victim's wife, and that of the victim. Each recounting describes a different motive and series of events. This leads to the phrase the *Rashomon effect*, meaning that what one sees depends upon where one sits.

## References

- Amabile, T.M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.
- Berliner, D. (1987). Expert and novice interpretations of classroom data. Paper presented at MidContinent Regional Education Laboratory Conference: Redesigning Schools, July 28–30, Breckenridge, CO.
- Bryan, W.L., & Harter, N. (1987). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, 4, 27–53.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Reprinted from *Handbook of Research on Teaching*. New York: Rand McNally & Company.
- Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- De Groot, A. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Leinmuntz (Ed.), *Problem solving*. New York: Wiley.
- Ford, M. (1987). An attitude survey of NASSP principal assessment candidates. Unpublished manuscript, Center for Educational Studies, Indiana University of Pennsylvania, Indiana, PA.
- Kintsch, W., & Van Kijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Lesgold, A.M. (1983, January). Acquiring expertise. (Technical report) Learning Research and Development Center, University of Pittsburgh.
- Newall, A. (1973). You can't play twenty questions with nature and win: Projective comments on the papers of this symposium. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Schmitt, N. Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. Criterion-related and content validity of the NASSP Assessment Center. Department of Psychology, Michigan State University, East Lansing, MI.

- Voss, J.F., Vesonder, G.T., & Spilich, G.J. (1980). Generation and recall by high-Knowledge and low-Knowledge individuals. *Journal of Verbal Learning and Verbal Behavior*, 19, 651-667.
- Voss, J.F., Greene, T.R., Post, T.A., & Panner, B.C. (1984). Problem solving skill in the social sciences. In G.H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.
- Wendel, F. (1988). *Implications of assessment center methodology for preparation programs in educational administration*. Arizona State University, UCEA Monograph Series.
- Wiedenbeck, S. (1985). Automation of programming skills. Report Series No. 2, Dept. of Computer Science, University of Nebraska, Lincoln, NE.