

# Student Performance: Assessment Centers Versus Traditional Classroom Evaluation Techniques

LYNN K. BARTELS

Southern Illinois University at Edwardsville  
Edwardsville, Illinois

WILLIAM H. BOMMER

Georgia State University  
Atlanta, Georgia

ROBERT S. RUBIN

St. Louis University  
St. Louis, Missouri

At business schools across the country, a movement toward skill-based learning is occurring in response to several factors: (a) students' inability to apply their business education to real-life business issues, (b) the emphasis placed by the American Assembly of Collegiate Schools of Business on assessing educational outcomes, and (c) employers' interest in hiring students with "good people skills." As a consequence of the focus on students' acquiring skills of this kind, new methods of assessing student performance are needed. Traditional assessment devices (e.g., multiple-choice tests, short answer or essay exams) may be unable to capture the new domain of interest. A few business schools have thus begun to use assessment centers to measure managerial skills. ACs use a variety of exercises to simulate managerial tasks. They offer "a higher fidelity" method of assessing learning than paper-and-pencil methods and may be better able to measure managerial skills (Riggio, Aguirre, Mayes, Belloli, & Kubiak, 1997). Additionally, assessment centers provide students with meaningful behavioral feedback regarding their strengths and weaknesses and recommendations for professional development.

Though ACs are new to business school curricula, they have been used for over 50 years by the military and pri-

**ABSTRACT.** Assessment center (AC) scores were compared with traditional large-class evaluation techniques (i.e., multiple-choice exams, group project grades, and discussion section scores), grade point average, and intelligence test scores. Undergraduate business students ( $n = 347$ ) were videotaped during an AC session consisting of an in-basket, two leaderless group discussions, and a speech. Trained raters used behavioral checklists to evaluate performance. AC scores correlated moderately with traditional evaluation techniques, suggesting that ACs measure knowledge acquisition in addition to skills acquired but not captured by traditional techniques. Their use is thus consistent with the emphasis on skill-based learning.

vate sector business (Thornton & Byham, 1982). As a result of their widespread use, ACs have broad acceptance among those who employ graduates of business schools (Spychalski, Quinones, Gaugler, & Pohley, 1997). Previous research has substantiated the reliability, validity, and fairness of assessment centers in predicting managerial success (Thornton & Byham, 1982).

Despite widespread acceptance of assessment center technology, their limited use in academic settings leaves a fundamental issue unresolved: Little research has assessed the relationships between traditional indicators of academic performance and performance in an assessment center. In the present study,

we examined the relationships between traditional measures of classroom performance, student ability, and performance in an academic assessment center designed to measure managerial skills.

## Grade Point Average

Grade point average (GPA) was thought to be a useful indicator because it represents the student's prior collegiate performance. GPA has been found to correlate with job performance according to meta-analytic research (corrected  $r$ 's in the .30's; Roth, BeVier, Switzer, & Schippmann, 1996) and has traditionally been the benchmark of academic success. However, GPA should not necessarily be thought of as a proxy for intelligence. Generally, intelligence does predict school grades well ( $r = .50$ , Neisser et al., 1996), yet GPA reflects many factors besides intelligence, including motivation, opportunity, interest, and persistence. Compared with more traditional forms of assessment (e.g., paper-and-pencil tests), assessment center exercises may bring out more of the aforementioned factors influencing GPA. For instance, delivering a speech to peers requires more than straight memorization and recall; it involves the abilities to synthesize and distinguish important information from the trivial and to express an idea clearly.

A central tenet of the skills-based learning model is that traditional academic outcome measures, which constitute a student's GPA, are deficient when it comes to evaluating skill acquisition. Thus, we examined the relationship between GPA and assessment center performance. The association of the traditional criteria (exam grades, discussion section grades, and group project grades) with AC scores was explored.

## **Cognitive Ability**

Cognitive ability has been used as a predictor of academic success (Neisser et al., 1996), and it may also play an important role in AC performance. The observed predictive validity of assessment centers may be attributable in part to the cognitive ability of participants. In the long-running AT&T Management Progress Study, Bray and Grant (1966) found that intellectual ability was an important contributor to overall performance. Klimoski and Brickner (1987) also suggested that intelligence may be one of the primary factors captured by ACs.

Cognitive ability plays a key role in predicting work performance. After completing an extensive review of personality and performance, Mann (1959) concluded that intelligence was the best predictor of performance in small groups. Small group behavior is often a significant part of AC performance. Since Mann's (1959) conclusions, numerous studies have corroborated the notion that intelligence is perhaps our best available performance predictor (Campbell, 1990; Hunter & Hunter, 1984; Ree & Earles, 1991). Hunter (1986) commented that a score from a cognitive ability test is as good a predictor of job performance as a composite score from tests measuring multiple abilities. Additionally, people who are higher in intelligence may be better able to manage the impressions of others. Huffcutt, Roth, and McDaniel (1996) found that during an interview, more intelligent job applicants were better able to convey a positive image of themselves than less intelligent applicants. Intelligence may create a similar advantage for assessment center participants. Consistent with the preceding argu-

ment, we hypothesized that cognitive ability would be significantly related to student AC performance.

Assessment centers may be used for administrative or developmental purposes, depending on the goals of assessment. When used administratively, ACs are used in decisionmaking on such things as promotion and grading. When used developmentally, the focus is on providing feedback to enhance skills. Many students and faculty believe that ACs should be solely developmental when used in an academic setting, in part because of the perception that AC methods are fundamentally different from traditional evaluation techniques and may alter the grading criteria substantially.

We argue that traditional assessment only measures the first stages of learning (i.e., declarative learning and knowledge compilation), whereas assessment centers capture all stages. According to Anderson's (1987) ACT\* model, learning occurs in three stages. In the first stage, declarative learning, learners acquire factual knowledge and information. Next, learners transform declarative knowledge in the knowledge compilation state (i.e., the learner moves from knowing "what" to knowing "how"). Finally, in proceduralization, learners apply and use knowledge to accomplish a task. According to this model, we would expect ACs to be correlated to traditional measures because both models require the same knowledge base for good performance. However, we would not expect a high correlation between ACs and traditional measures, because of the absence of behavioral application in traditional measures. Therefore, we argue that ACs offer a measurement tool that captures all stages of the learning process and can account for true managerial skill learning beyond simple knowledge acquisition. Thus, students who perform well on traditional measures, but not in an assessment center, may not have fully learned the necessary managerial skills.

## **Method**

### *Participants*

Undergraduate business students ( $n = 347$ ) from a large midwestern university

participated in a managerial assessment center as an organizational behavior course requirement. The students could best be described as traditional undergraduates at a residential Carnegie Research 1 institution. The sample was 45% female, and about 85% classified themselves as "White/Non-Hispanic."

### *Assessment Center Procedures*

Participants assumed the role of department manager in a computer company. Background information about the company was provided 1 week in advance. During the 150-minute assessment, participants attended two 20-minute leaderless group discussion meetings (budget and selection), delivered a speech, and worked on an in-basket.

At the assessment center, participants received an in-basket with additional organizational information, along with a stack of materials including memorandums, appointments, and so forth. In the budget meeting, seven participants worked together to develop a budget for the upcoming fiscal year. Each participant had information regarding his or her department's needs. In the selection meeting, the participants identified potential successors for the CEO of their company. Each participant had seven resumes from CEO candidates, and the group was asked to rank the top three candidates. In addition, each participant presented a 3-minute speech highlighting his or her vision for the organization. The speech was to be developed at the assessment center using materials from the in-basket.

### *Measures*

*Assessment center performance.* The meetings and speeches were videotaped, and trained raters assessed participants' performance. Frame-of-reference rater training (Bernardin & Buckley, 1981) consisted of a 2-day workshop designed to help the raters become proficient with the AC materials. Multiple videotaped examples were used for practice and raters were tested for certification at the training's conclusion.

Four assessment dimensions—oral communication, teamwork, future ori-

**TABLE 1. Means, Standard Deviations, and Intercorrelations**

| Variables             | <i>M</i> | <i>SD</i> | 1     | 2     | 3     | 4    | 5     | 6     | 7     | 8   | 9   |
|-----------------------|----------|-----------|-------|-------|-------|------|-------|-------|-------|-----|-----|
| 1. Overall AC score   | 462.03   | 95.57     |       |       |       |      |       |       |       |     |     |
| 2. In-basket          | 106.81   | 36.05     | .36** |       |       |      |       |       |       |     |     |
| 3. Selection meeting  | 136.81   | 59.68     | .73** | -.04  |       |      |       |       |       |     |     |
| 4. Budget meeting     | 132.47   | 41.38     | .60** | .01   | .25** |      |       |       |       |     |     |
| 5. Speech             | 85.09    | 32.41     | .44** | .07   | .09   | .04  |       |       |       |     |     |
| 6. GPA                | 3.08     | .40       | .23** | .21** | .11*  | .03  | .25** |       |       |     |     |
| 7. Exams              | 83.23    | 8.44      | .24** | .24** | .14*  | .04  | .19** | .56** |       |     |     |
| 8. Group project      | 91.33    | 3.61      | .07   | .02   | .04   | -.02 | .14** | .09   | .06   |     |     |
| 9. Discussion section | 90.64    | 5.72      | .13*  | .04   | .12*  | -.02 | .17** | .15** | .15** | .05 |     |
| 10. Cognitive ability | 25.03    | 4.73      | .24** | .26** | .10   | .13* | .13*  | .22** | .16** | .06 | .02 |

Note. *N* ranges from 327 to 347.

\* $p < .05$ .

\*\* $p < .01$ .

entation, and time management—were assessed using behavior checklists (Reilly, Henry, & Smither, 1990). Sixty managers were asked to rate the importance to managerial success of each behavior from the checklist. The ratings were then used to weight each behavior and each behavior dimension's importance to the exercise score. An overall performance measure was computed as a weighted sum of the scores from the four exercises.

**Grades.** Participants self-reported their current grade point averages on the demographic questionnaire administered at the end of the AC evaluation. Multiple-choice exams were augmented by two additional grading opportunities. First, students were assigned to groups to work on a project in which they proposed solutions to real-life problems of area businesses. This project culminated in presentation of the group's recommendations to the company. In addition, the large lectures were divided into groups of 20–25 students for once-a-week discussion sections. The grade from these discussion sections was based on attendance, participation, and small writing assignments.

**Intelligence.** Intelligence was measured using the Wonderlic Personnel Test (1992). The 12-minute test consists of 50 items, including verbal ability problems and quantitative/logic ability items. Created as a tool for use in employment selection, the Wonderlic

(1992) has shown high predictive validity in measuring general cognitive ability. For instance, it is highly correlated with other well-known measures of intelligence (e.g., WAIS-R,  $r = .96$ ; GATB,  $r = .80$ ). Additionally, the Wonderlic (1992) has demonstrated strong test-retest reliability estimates ranging from .82 to .94, as well as high internal consistency ranging from .88 to .94. The Wonderlic was administered at the completion of the AC process.

## Results

Means, standard deviations, and intercorrelations among the variables are presented in Table 1. Examination of the intercorrelations shows that GPA was significantly related to overall AC scores ( $r(337) = .23, p < .01$ ). Overall AC scores also correlated with the other grading criteria. Those students who performed better in the AC also scored higher on the multiple-choice exams ( $r(344) = .24, p < .01$ ) and received higher discussion section grades ( $r(344) = .13, p < .05$ ). There was, however, no relationship between AC performance and group project grades ( $r(344) = .07, ns$ ).

The data in Table 1 show that cognitive ability was significantly correlated with overall AC performance ( $r(344) = .24, p < .01$ ). Students who scored higher on the cognitive ability test tended to receive higher AC scores. Cognitive ability significantly correlated with three of the four exercises that made up

the AC (cognitive ability/in-basket  $r(332) = .26, p < .01$ ; cognitive ability/selection  $r(332) = .10, ns$ ; cognitive ability/budget  $r(332) = .13, p < .05$ ; cognitive ability/speech  $r(332) = .13, p < .05$ ).

## Discussion

The purpose of this study was to investigate the relationship between AC performance and other student evaluation techniques. Results indicated that GPA, course exam grades, and discussion section grades were all correlated with AC performance. The only grading criterion that was not related to AC performance was the group project grade. This is likely because of the restricted range for the group project grades; also, the group project grades may be more of a reflection of group performance than individual performance. Assessment center ratings, on the other hand, were based on individual performance, even during group meetings. There was, however, a correlation between group project grades and speech ratings, which might be due to the oral presentation component of the group project.

As predicted, cognitive ability and AC performance were correlated. More intelligent students tended to perform better in the AC. Cognitive ability was a particularly good predictor of in-basket performance. Thus, for the exercise that included written communication and no social interaction, cognitive ability was a good predictor. This finding is consis-

tent with Shore, Thornton, and Shore (1990), who found that cognitive ability correlated more strongly with task-oriented dimensions than interpersonal-style dimensions. Similarly, Goldstein, Yusko, Braverman, Smith, and Chung (1998) found that Wesman Personnel Classification test scores, which measure intelligence, were most strongly correlated with the in-basket exercise scores. They concluded that intelligence was more strongly related to scores in AC exercises that have a higher cognitive loading.

Although GPA was related to overall AC performance, and performance in most exercises, it tended to predict individually focused performance (e.g., in-basket and speech scores) best. These are the exercises that are probably most similar to the types of assessments included in GPA (paper-and-pencil tests, papers, presentations, cases, etc.). In spite of efforts to incorporate group exercises into academic coursework, GPA appears to be more reflective of individually based performance than group-based performance.

#### *Implications for Business Educators*

Academic ACs show promise in helping educators narrow the gap between knowledge and application. In this endeavor, ACs may serve developmental as well as administrative purposes. The present data suggest that ACs tap a domain consistent with the knowledge typically assessed by traditional measures. In addition, ACs may also go beyond traditional assessments in providing a higher level assessment of skill acquisition. Thus, it would seem appro-

priate not only to provide developmental feedback but also to base course grades on AC outcomes.

If business school educators are serious about producing highly skilled graduates, this study suggests that our traditional outcome measure (i.e., exam scores) is a useful starting point, but by no means a complete indicator of managerial skill learning. The present study suggests that many managerial skills may be insufficiently assessed in business school curricula. Thus, we need to improve our assessment techniques if we are really concerned about assessing skill-based outcomes.

#### NOTE

A version of this article was presented at the Midwest Academy of Management Conference, 18 April 1998, Kansas City, Missouri.

#### REFERENCES

- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, *94*, 192-210.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review*, *6*, 205-212.
- Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, *80*(17), 1-27.
- Campbell, J. P. (1990). An overview of the army selection and classification project (project A). *Personnel Psychology*, *43*, 243-257.
- Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercise. *Personnel Psychology*, *51*(2), 357-374.
- Hufcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, *81*(5), 459-473.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340-362.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-88.
- Klimoski, R., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, *40*, 243-260.
- Mann, R. D. (1959). A review of the relationships between personality and performance in small groups. *Psychological Bulletin*, *56*(4), 241-270.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77-101.
- Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*, 321-331.
- Reilly, R. R., Henry, S., & Smither, J. W. (1990). An examination of the effects of using behavioral checklists on the construct validity of assessment center dimensions. *Personnel Psychology*, *43*, 71-84.
- Riggio, R. E., Aguirre, M., Mayes, B. T., Belloli, C., & Kubiak, C. (1997). The use of assessment center methods for student outcome assessment. In R. E. Riggio & B. T. Mayes (Eds.), *Assessment centers: Research and applications* [Special Issue]. *Journal of Social Behavior and Personality*, *12*(5), 273-288.
- Roth, P. L., BeVier, C. A., Switzer, F. S., & Schippman, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, *81*(5), 548-556.
- Shore, T. H., Thornton, G. C., III, & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology*, *43*, 101-115.
- Spychalski, A. C., Quinones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, *50*(1), 71-90.
- Sternberg, R. J. (1997). Does the graduate record examination predict meaningful success in the graduate training of psychologists? *American Psychologist*, *52*(6), 630-641.
- Thornton, G. C., & Byham, W. C. (1982). *Assessment centers and managerial performance*. San Diego, CA: Academic Press.
- Wonderlic Personnel Test, Inc. (1992). *The Wonderlic Personnel Test manual*. Liberty, IL: Association of Test Publishers.