DOCUMENT RESUME

ED 351 361   TM 019 173

AUTHOR   Jacobson, Larry; Stilley, Lori Rittenhouse
TITLE   Developing and Scoring Assessment Center
                Exercises.
PUB DATE   Apr 92
NOTE   14p.; Paper presented at the Annual Meeting of the
                National Council on Measurement in Education (San
                Francisco, CA, April 21-23, 1992).
PUB TYPE   Reports - Descriptive (141) -- Speeches/Conference
                Papers (150)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Assessment Centers (Personnel); College Students;
                Elementary School Teachers; Elementary Secondary
                Education; Evaluators; Higher Education; *Licensing
                Examinations (Professions); Pilot Projects; *Scoring;
                Secondary School Teachers; Simulation; Site Analysis;
                Situational Tests; *Teacher Certification; Teacher
                Education; Teacher Evaluation; *Test Construction
IDENTIFIERS   Subject Content Knowledge; Teacher Candidates;
                *Teacher Competency Testing

ABSTRACT
        The Assessment Development Laboratory for the
National Board for Professional Teaching Standards is developing a
certification process with the three components of school site
documentation, a content knowledge examination, and assessment center
exercises. Activities involved in the development of the exercises
and their scoring are described. The assessment center method is a
process that includes: (1) standardized assessments based on multiple
scores of candidate evidence; (2) multiple trained assessors; (3)
judgments about evidence based, in part, on simulation exercises; and
(4) judgments pooled by assessors or by statistical integration. A
major consideration in developing the exercises was the
representation of the exercises to teaching. Six dimensions of
teacher tasks were identified, and each exercise was designed to
elicit evidence on at least three dimensions. Exercises were also
designed to elicit pedagogical reasoning and action. Resource needs
for assessment administration were a further consideration. Four
exercises are currently being developed, centering on cooperative
group discussion, instructional analysis, planning instruction, and
evaluating student learning. Small pilot tests (smoke tests) have
been administered in the development process. A detailed analysis of
each exercise illustrates the development process. (SLD)

DEVELOPING AND SCORING
ASSESSMENT CENTER EXERCISES

Larry Jacobson
Connecticut State Department of Education

Lori Rittenhouse Stilley
University of Pittsburgh

2

INTRODUCTION

The Assessment Development Laboratory (ADL) is developing a certification process that has three components: school site documentation, a content knowledge exam, and assessment center exercises. In this presentation we will describe the activities, to date, involved in the development of exercises and scoring within the assessment center component of the ADL.

Initially the ADL adopted the assessment center method as defined by the International Congress on the Assessment Center Method in *Guidelines and Ethical Consideration For Assessment Center Operations* (1989).

The assessment center method can be described as a process that includes; (a) standardized assessments based on multiple sources of candidate evidence; (b) multiple trained assessors; (c) judgments about evidence based, in part, on simulation exercises; and (d) judgments pooled by assessors or by statistical integration.

Although all three of the components within the ADL certification process fall within the domain of an assessment center, for purposes of this presentation, the phrase, *Assessment Center*, will be used to specifically refer to the third component of the ADL. Currently, four simulation exercises are being considered for the assessment center component.

DEVELOPMENTAL FRAMEWORK OF EXERCISES

For over 40 years, assessment centers have been conducted in the United States for management and supervisory professions (Howard and Bray, 1988). During that time a variety of exercises such as leaderless group discussions, in-baskets, oral presentations, budget and planning activities have evolved. These exercise formats are well known and have been developed to assist in predicting successful managerial performance. By contrast, the Education profession has only begun to develop exercises best suited to the assessment of classroom teachers. It is anticipated that common exercise types, just as they have evolved in other professions, will form for assessment of early adolescent English language arts teachers. In contrast, however, to some of the traditional applications of the assessment center method, exercises developed for the National Board have been specific to the subject area and grade level of the certification. As an illustration, consider a management in-basket exercise which typically asks candidates to make decisions in response to day-to-day tasks (e.g., memos, requests from colleagues and administrators, and the like). Although such an exercise could be adapted to elicit some of the knowledge, skills and abilities required of classroom teachers, the exercise format itself, based on

experiences of the ADL, would not be well accepted because handling such paperwork, although necessary in teaching, is not sufficient to represent the Standards which define EA/ELA accomplished teaching.

Another key premise of the assessment center component is that simulations can be designed to evoke many of the knowledge, skills and abilities underlying accomplished teaching. Furthermore, the assessment center component has, as an advantage over other methods such as direct observation of teaching, the opportunity to compare candidate performance across a common set of stimuli and to examine candidate reasoning utilizing face-to-face interviews. For example, an interviewer can probe within specified guidelines for clarification of candidate's reasoning related to particular actions or decisions.

Exercise Guidelines

Four considerations have guided the ADL in the development of the assessment center component. These considerations include the representation of the e _rcises to teaching, dimensional evidence expected to be elicited by the exercise, how exercises elicit pedagogical reasoning and action, and the resources required to conduct the exercise.

Representation of the Exercise to Teaching

The ADL began the developmental process by examining the Early Adolescence/English language arts Standards and the ADL Dimensions to find critical aspects of teaching that could be represented as standardized exercises and that would focus on the knowledge, skills and abilities thought to generalize across the ADL certification components. The assessment center tasks were designed to simulate these critical aspects of English language arts teaching and have been selected to be representative of the key tasks that accomplished teachers perform. In addition, exercises were developed to represent the complexity and context of teaching in uniform format. For purposes of administrative feasibility, a balance was reached between exercise complexity and realism with the recognition of what candidates can be expected to accomplish within a two to three hour exercise. Exercise types currently include, discussion of curricular issues, analyzing instruction, planning coherent instruction, and evaluating student writing.

As previously discussed, the assessment center component examines candidate performance from a perspective that differs from the school site portfolio component in which teachers' performances are situated within the context of their own classrooms. When designing the exercises, tasks were selected that could be standardized but still would appear realistic to candidates. Some exercise types (e.g., simulations of teaching with role-players or actual students) were dismissed due to issues of appropriateness and feasibility. For example, asking candidates to teach students with whom they had no previous experience represented an unrealistic expectation of candidates. Furthermore,

training and logistics involved in the use of role-players and/or students posed a variety of administrative problems.

It was also important that the exercise be administered and scored within a reasonable time frame. Again, a balance was created between the complexity of the exercise and the demands placed on the interviewers and judges. For example, when asking candidate's to evaluate a set of student papers, it might appear more "authentic" to candidates to bring in their own student papers to the interview. Doing so, however, would require each interviewer and judge to become familiar with different sets of papers and, more troublesome, would no longer provide judges with a common frame of reference in which to assess candidates. By contrast, developing an exercise scenario that asks candidates to evaluate a standardized set of papers written by students at the end of their previous year, not only makes the assessment process more manageable and meaningful to candidates.

Dimensional Evidence

As discussed earlier in this symposium, the ADL dimensions, which are derived from the National Board Propositions and Early Adolescence/English Language Arts Standards, served as the foundation upon which exercises have been constructed. (See Pence and Petrosky, 1992, for more information on the standards development process and its relationship to exercise development). The ADL dimensions provide the framework for designing assessment center exercises and for assessing the candidate's performance on an exercise. During development, exercises had to represent the types of tasks that EA/ELA teachers should know and be able to do, but also had to elicit dimensional evidence that would likely result in a range of candidate performance.

The six ADL dimensions are as follows:

A. Teachers understand and respond to students' knowledge, beliefs, attitudes, and interests. (Knowledge of Students)

B. Teachers understand and respond to the nature of cultural diversity in literature, language, and society (including the classroom). (Cultural Diversity)

C. Teachers understand the diverse aspects of English language arts and the interrelationships among its various aspects. (Content Knowledge)

D. Teachers understand and use an integrated approach to the teaching of English language arts. (Integrated Pedagogy)

5

E. Teachers understand and use a coherent pedagogy in the teaching of English language arts. (Coherent Pedagogy)

F. Teachers understand and respond to professional concerns in English language arts. (Members of Learning Community)

Each exercise was designed to elicit evidence on three or more of the dimensions.

Pedagogical Reasoning and Action

Exercises were also designed to elicit what Shulman (1986) defines as pedagogical reasoning and action. Exercises were developed to elicit candidates' reasoning, their demonstration of knowing the content of instruction and how they think about their own teaching practices; and/or to elicit candidates' action, demonstrating that they know how to teach. During an exercise, candidates provide evidence of a particular dimension through their actions and/or responses during the interview.

Working from Shulman's pedagogical reasoning and action, the ADL tried to achieve a balance of a teacher's knowledge, application of knowledge, and reflection across all three components of the certification process. The assessment center exercises were particularly useful in tapping the teacher's reasoning and reflection.

Resource Needs

The feasibility of conducting exercises that can represent some of the complexities involved in teaching has required careful attention. Some of the significant resources needed to administer a large-scale assessment center component include, appropriate examining facilities; audio and video recording equipment; exercise administrators, interviewers and judges; and resources associated with monitor, interviewer and judge training. Of special note is the necessity of having subject matter experts (EA/ELA classroom teachers) serve as interviewers and judges. The requirement that interviewers and judges possess similar content-pedagogical experiences to that of the candidate group has placed additional constraints on the availability of examiners.

EXERCISE FORMATS

In preparation for the larger-scale field test, four exercises are currently being developed. The assessment center component is expected to require from one and one-half to two days of candidate time. The general framework for each exercise first provides candidates with an opportunity to read and prepare for a

group discussion or to prepare for a semi-structured interview. The following exercise types are being considered and/or being smoke-tested:

- Cooperative Group Discussion Exercise

The cooperative group discussion asks candidates to discuss a curricular issue with three other colleagues (all of whom are candidates) and make a group recommendation to a posed problem. This exercise is designed to focus on the candidates' knowledge of selected issues and their ability to make recommendations based on this knowledge while demonstrating their abilities to work collaboratively with fellow professionals in the field. Prior to the group discussion, candidates are provided time to consider the issue and prepare recommendations. Then, as a group, candidates discuss the issue and attempt to reach consensus on recommendations.

- Instructional Analysis Exercise

The instructional analysis exercise asks candidates to analyze a segment of another teacher's instruction by explaining the strengths and weaknesses and then making recommendations for improvement. For example, the candidate might be asked to review a "teacher's" goals; activities, integration of reading, writing, speaking, and listening; or evaluation of students. The candidate would read some background information provided by the "teacher" about her classroom instruction and would then view a videotape of a short example of her instruction. The candidate is provided time to analyze the instruction and then asked about their analysis in a semi-structured interview.

- Planning Instruction

This exercise is designed to focus on the candidate's understanding of an English language arts concept and his or her ability to translate this knowledge into a coherent plan for instruction. The planning exercise, for example, might ask candidates to plan a coherent segment of instruction, possibly two to three days, on the concept of language variation. The particular concept of English language teaching is selected to represent a major topic in English language arts and one with which early adolescent/English language arts teachers should be familiar. The candidate would be provided time to plan instruction and then, in a semi-structured interview, asked to explain his or her instructional decisions.

- Evaluating Student Learning

This exercise is designed to simulate a teacher's assessment of a range of abilities within a group of students' writing and focuses on the candidate's ability to develop and apply criteria for evaluating students' writing. This exercise

differs from the school site student learning exercise in several ways. This exercise is not contextualized to the candidate's classroom and instead uses a standardized set of student papers, focuses on a variety of students, and asks the candidate to look at summative versus formative written assessment.

SMOKE-TEST PURPOSE AND PROCESS

In order to prepare for the larger-scale field test to take place in the fall, a series of small group pilots were conducted. These small-scale pilots have been referred to as smoke-tests and were conducted for the following purposes:

- To collect information to guide the revision of exercises (i.e., increase the opportunity to elicit dimension evidence).

- As an opportunity for EA/ELA teachers to examine the degree to which the exercises accurately represent tasks that teachers need to know and be able do.

- To examine the effects of the exercises on a variety of candidates (e.g., diverse grade levels, gender, and ethnicity).

- To refine candidate, interviewer, and monitor directions.

- To record candidate performance to assist in the development of the interviewer and judging workshops.

To date a series of smoke-tests have been administered and have been recorded on audio and/or videotape. In addition, transcripts have been made for review and refinement of the scoring process, and all smoke-tests have culminated in group feedback sessions, with teacher participants asked to provide feedback to improve the exercises.

Cooperative Group Discussion Exercise

A description of the development of the cooperative group discussion exercise will serve to illustrate how the four considerations (the representation of the exercises to teaching, dimensional evidence expected to be elicited by the exercise, how exercises elicit pedagogical reasoning and action, and the resources required to conduct the exercise) impact the assessment center exercise development process.

Representation of the Exercise to Teaching

The cooperative group discussion exercise has been designed to examine how an EA/ELA teacher will work with other teachers when confronting professional issues. One of the first goals involved in developing this exercise was to identify discussion issues that represent some of the challenges facing EA/ELA teachers and that represent important issues that are, or are likely to be, discussed by teachers. After considering a variety of possible discussion issues, the area of curriculum selection was chosen, as it has been found to be a critical professional activity and ideally suited to this exercise format.

Another goal in the development of this exercise was to simulate a collegial group discussion by allowing the group discussion to proceed with as little structure as possible. Earlier, the Stanford Teacher Assessment Project (Athanases, 1990), designed a group exercise that attempted to structure the group discussion to ensure, among other concerns, that each candidate had equal opportunity to speak. This approach changed the discussion into a structured recitation, and TAP researchers suggested that this type of discussion exercise may not be viable. Consequently, the ADL was faced with the question: Is it possible to allow candidates to have a group discussion without monitor intervention yet provide sufficient focus to allow for an evaluation of dimensions?

The ADL considered the potential advantages and disadvantages that might result during an unstructured group discussion, and proceeded to develop a process designed to address candidate equity and scorability of the exercise. The ADL addressed two key areas to facilitate the group discussion process: issue familiarity and group discussion guidelines.

Content familiarity

Since the focus of this exercise is on a participant discussion and less on a candidate's knowledge of specific curriculum materials, there was initial concern about the range of the participants' prior content familiarity. If a candidate has little or no familiarity with a particular text used in the exercise then participation during the discussion could be affected. To assure that participants

came into the exercise with similar preparation, examinees received copies of text materials approximately 4-6 weeks in advance. This advanced notice allowed candidates an opportunity to become familiar with the materials as they found necessary.

Group discussion guidelines

Immediately before the group discussion participants are provided with several administrative guidelines such as not assigning a group recorder or leader, and limiting discussion to the texts within the exercise and time constraints.

Each of these pre-discussion procedures were built into the exercise to ensure that the group discussion could proceed naturally, but with sufficient focus so that the participants did not stray from the goals of the task.

Dimensional Evidence

The CGD exercise was developed with the primary goal of eliciting candidate evidence associated with certain dimensions. This exercise is designed to elicit evidence on five of the six ADL dimensions: knowledge of students (A), cultural diversity (B), content knowledge (C), integrated pedagogy (D), and responding to professional concerns (F). While discussing a curricular issue, teachers are asked to focus on their knowledge of students, the cultural diversity of the instruction, their knowledge of instruction, and how they would integrate the instruction, while demonstrating their ability to work collaboratively with peers.

To increase the likelihood that candidates focus their discussion on the pertinent dimensions, they were first provided with a set of considerations, which refer to key aspects of the dimensions elicited by the exercise, to guide their preparation for the discussion. In addition, prior to the group discussion, participants are asked to refer to these considerations during their discussion.

Pedagogical Reasoning and Action

The cooperative group discussion was designed to elicit the candidate's reasoning through the action of a group discussion. An exercise scenario was designed where an assistant superintendent asks a small group of English teachers for their recommendations on selecting curriculum materials. Through the action of the discussion, the candidates demonstrate their knowledge of curriculum selection as it relates to students and instruction, and provide rationale for making particular curricular recommendations.

Resources Required

Initially, the cooperative group discussion exercise involved more resources than other exercises due to the addition of individual "debriefing" sessions. A review of audio and videotapes of the smoke-tests has suggested that individual debriefings can be effectively conducted in a written format. Consequently, the cooperative group discussion exercise is expected to require fewer resources than other assessment center exercises.

Conclusions

Based on the results from a series of smoke-tests conducted in Pennsylvania and Connecticut, the ADL has concluded that the cooperative group discussion exercise can be both efficient and effective. The cooperative group discussion exercise requires fewer administrative and scoring resources than exercises utilizing a semi-structured interview. In addition, the discussion exercise provides a sample of candidate performance not observable within any other component/exercise. Further, smoke-test evidence suggests that the group exercise can elicit candidate evidence from a variety of dimensions and has yielded an additional benefit of high face validity based on enthusiastic participant endorsements of this exercise. It is also important to note that many of the participants have expressed an interest in doing more cooperative group work in their own schools based on their experiences with this exercise.

DEVELOPMENT OF A SCORING PROCESS

As described earlier in the session (Delandshere and Pecheone, 1992), scoring has been integrated in all phases of the development process. For example, each exercise type was selected based on its potential to elicit dimension evidence from candidates. As smoke-test tapes and transcripts were reviewed, participant performance was examined with respect to dimensions and exercises and dimensions wei "fine-tuned." Accordingly, although small-scale, the smoke-tests have provided a means of revising the exercises based on areas where dimension evidence appears to be too thin or non-existent. For example, if an exercise does not appear to reveal the dimensional evidence of interest, the exercise may be revised to better elicit such evidence.

Structure of the Scoring Process - Introduction

Typically, the assessment center method involves a "live" scoring process (Thornton and Byham, 1982). For example, during the group discussion exercise, assessors observe assigned candidates, then review notes and related materials, and conclude with a rating of the candidate's performance on dimensions related to this exercise. This live scoring process requires, among other things, sufficient

prior training to ensure that assessors understand and uniformly apply ratings according to certain benchmarks or exemplars of candidate performance. Since, the assessment center component will not, as of yet, have a sufficient range of candidate performance on which to train and calibrate judges prior to the field test, the ADL will separate the field-test into an administrative and a scoring phase. In the administrative phase, all exercises will be administered and participant performance will be video and audiotaped. In the scoring phase, conducted later, videotapes will be selected for purposes of judge training and calibration, and judges will then be convened to evaluate the recorded candidate performance. Following the field-test, it is anticipated that the administrative and scoring phases may be combined as a more efficient assessment process.

Cooperative Group Discussion Scoring

The cooperative group discussion exercise will serve as an example of how an assessment center exercise may be scored. The ADL has developed a high-inference scoring system based on the six dimensions (see Delandshere and Petrosky, 1992). The system uses pairs of expert judges to evaluate and rate exercises. The process begins when the judges observe the videotape of the cooperative group discussion (see Figure 1). Judges then review notes and consider questions that are based on the ADL dimensions and the exercise. For example, for dimension A (Knowledge of Students) questions might be as follows:

How does the teacher take into account the students' interests, backgrounds, and experiences when selecting curriculum materials?

How does the teacher take into account the students' need and abilities when selecting curriculum materials?

As judges consider candidate evidence, guides which are customized to each exercise are provided to further define the questions by focusing on critical facets of each dimension. After taking notes, the judges discuss the candidate's performance and begin to characterize the evidence within a given dimension. The pair writes an interpretive summary of the evidence for each dimension and then rates that performance on a four-point scale that has been anchored to prejuried exemplars (i.e., videotapes and interpretive summaries that have been anchored to the rating scale). The pair also evaluates the confidence that they have in their rating based on the quality or sufficiency of the available evidence.

CONCLUSION

The assessment center component proposed for the National Board for Professional Teaching Standards (Early Adolescence/English Language Arts) has been designed to elicit key knowledge, skills, and abilities thought to underlie accomplished teaching. Through the use of exercises which strike a balance

between standardized simulations and realistic situations, participant performance will be recorded and later evaluated by judges who possess similar content-pedagogical backgrounds to candidates being assessed, and have been trained and calibrated to uniformly apply standards. Each of the steps described in this process has been taken to provide sources of information in which to make judgments about the reliability, validity, and administrative feasibility of the assessment center component.

References

Athanases, S. Z. (1990) Assessing teacher skills and knowledge in the selection and use of literature in the elementary grades. (Report No. L5). The Teacher Assessment Project: Stanford University.

Delandshere, G. and Petrosky, A. (1992). Developing a high inference scheme to evaluate teacher knowledge. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco.

Delandshere, G. and Pecheone, R. (1992). Dimensions: Basis for an integrated assessment. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco.

Guidelines and ethical considerations for assessment center operations (1989). Task Force on Assessment Center Guidelines, *Monograph XVI*, Development Dimensions International.

Howard, A., and Bray, D. (1988). *Managerial lives in transition.* New York: The Guilford Press.

Pence, P. and Petrosky, A. (1992). Defining performance standards and developing assessment exercises for accomplished English language arts teaching of young adolescents. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher.* 15(1). 4-14.

Thornton, G.C. and Byham, W. C. (1982). *Assessment Centers and Managerial Performance.* California: Academic Press.