

2020

## Behavioral Assessment of Expert Talent Competencies: Analysis, Decision Making, and Written and Verbal Communication Skills

Homayoon Kord  
*Mobarakeh Steel Company, homayoon.kord@gmail.com*

George C. Thornton III  
*Colorado State University, george.thornton@colostate.edu*

Follow this and additional works at: <https://scholarworks.bgsu.edu/pad>

 Part of the [Human Resources Management Commons](#), and the [Industrial and Organizational Psychology Commons](#)

### Recommended Citation

Kord, Homayoon and Thornton, George C. III (2020) "Behavioral Assessment of Expert Talent Competencies: Analysis, Decision Making, and Written and Verbal Communication Skills," *Personnel Assessment and Decisions*: Vol. 6 : Iss. 1 , Article 4.

DOI: <https://doi.org/10.25035/pad.2020.01.004>

Available at: <https://scholarworks.bgsu.edu/pad/vol6/iss1/4>

This Measurement and Measures is brought to you for free and open access by the Journals at ScholarWorks@BGSU. It has been accepted for inclusion in Personnel Assessment and Decisions by an authorized editor of ScholarWorks@BGSU.

---

## Behavioral Assessment of Expert Talent Competencies: Analysis, Decision Making, and Written and Verbal Communication Skills

### Erratum

A previous version of this article included incorrect values in Table 5. These values were corrected and a small change to the text was made. Correction of these errors did not change any conclusions reported in the previous version.

# BEHAVIORAL ASSESSMENT OF EXPERT TALENT COMPETENCIES: ANALYSIS, DECISION MAKING, AND WRITTEN AND VERBAL COMMUNICATION SKILLS

Homayoon Kord<sup>1</sup> and George C. Thornton III<sup>2</sup>

1. Mobarakeh Steel Company

2. Colorado State University

Errata: A previous version of this article included incorrect values in Table 5. These values were corrected and a small change to the text was made. Correction of these errors did not change any conclusions reported in the previous version.

## ABSTRACT

### KEYWORDS

expert competency,  
behavioral assessment,  
assessment center

Organizations face challenges of screening applicants for critical skills to serve in expert staff positions requiring interactions with line managers. Such positions require a special set of cognitive and interpersonal competencies. This study investigates the psychometric qualities of a new behavioral assessment method in use in an applied setting. Using data from a group of 219 finalists for positions in a large Iranian steel company, it examined the validity and fairness of the method in relation to other test and demographic information. Results showed evidence of convergent and discriminant validity and no discrimination against women or older candidates. The study contributes to a clearer understanding of expert competencies and a practical method for assessing and training such competencies. Next steps and future needs are identified.

Competencies comprising talent such as inductive reasoning, deductive reasoning, and oral and written comprehension and expression, are needed by organizational staff members serving line managers. They are important for many jobs (ONET, <https://www.onetcenter.org/content.html>) and central components of models of leadership (Church & Silzer, 2014; Paschen & Dihmaier, 2014; Thornton & Reynolds, in press). Past efforts to assess these competencies have been partially successful, but more comprehensive and versatile methods are needed. Therefore we describe a new method in operation to assess talent competencies and report research into its psychometric qualities. The purpose of this study is to develop and evaluate a new behavioral assessment method, Analytical Writing and Discussion (AWD) to assess talent competencies.

We organize this article in the following sections. First, we review the importance of talent competencies, alternatively named skills or dimensions. Next we review past efforts to assess talent competencies. Then we describe AWD, a new measure to assess such skills. The method to evaluate AWD involved the first application of the measure to a large sample of employees in a large Iranian organization who were given the AWD and other measures. Next we present evidence of the quality of AWD, including psychometric analyses of validity. We conclude with a summary of strengths and limitations, along with recommendations for future applications and research needs.

## Expert Talent Tasks and Competences

Experts in organizations have special assignments and carry out special tasks. They study, analyze, give suggestions, and provide alternatives for managers to make decisions and plans. They collaborate with managers in designing, running, and controlling processes. In addition, they develop procedures, prepare instructions, and train operators to carry out procedures.

Experts are often in staff positions in contrast with line positions occupied by what are known as supervisors, department heads, and executives. Line managers have positions with formal authority by virtue of their position in the hierarchy of the organization. By contrast, staff experts typically do not have formal authority, but rather, they have informal authority as they give advice to line managers. Their tasks include writing reports and discussing recommendations with managers. Because of these differences from line managers, staff experts have to rely on different competencies in comparison with the competencies of formal leaders. For example, in many modern organizations

Corresponding author:  
George C. Thornton  
Department of Psychology Colorado State University Fort Collins,  
CO 80524  
Email: [george.thornton@colostate.edu](mailto:george.thornton@colostate.edu)

experts do their tasks through teamwork, especially in teams without formally designated leaders. They may have to contribute, collaborate, and lead a diverse team in order to develop a new product or a technology.

Expert talent in organizations is the complex multifaceted ability to understand the needs of colleagues, provide relevant advice, and persuade others of the rationale and value of this advice, all without formal organizational authority. Such talent requires not just knowledge and skills associated with a specialty professional discipline such as engineering, marketing, or information technology but also cognitive abilities to understand the conceptual problems in a diversity of jobs and interpersonal competencies to communicate with others and persuade others that recommendations and decisions are valid.

Exercising expertise is a form of a leadership as described by *Paschen and Dihmaier (2014)*, who emphasize that, in essence, leadership is exercise of power through persuasion. This form of leadership can be exercised at all levels of organizations, not only in management positions that have formal power but also in expert positions where staff members provide expert advice to managers. Experts seeking to influence others may need even more leadership skill than managers because persuading others without formal organizational authority is often much more difficult. Furthermore, the need for experts to demonstrate informal leadership is increasingly important in the current era as organizations downsize, decrease levels of line management, structure work into leaderless group, and expect active participation from all levels of employees. As a consequence, broader empowerment is required throughout the organization.

Methods of assessing talent are useful for several purposes: screening and selecting personnel, diagnosing strengths and weaknesses in components of expert talent, training experts in their roles, and evaluating effectiveness of expert training programs. Measures of talent can be used in combination with measures of other competencies such as cognitive ability and leadership. Measures of pure cognitive ability have been established to be one of the predominant competencies essential for effectiveness in any position in an organization (*Schmidt & Hunter, 1998*). Scores of models of leadership have been formulated from a theoretical or practical perspective (*Day, Fleenor, Atwater, Sturm, & McKee, 2014*). These models virtually all include some combination of cognitive and interpersonal competencies. Thus, AWD was designed to assess a range of behavioral competencies that complement measures of related competencies.

### Methods of Assessing Expert Competencies

Methods of assessing expert talent are different from methods of assessing managerial competencies such as

planning and organizing, delegation, and even formal leadership. They must be applicable to individuals from different disciplines and organizational departments, and to individuals providing advice to managers at different levels in the organization. A sample of commonly used assessment methods includes the following, each with strengths and weaknesses.

### Psychometric Tests

The strengths of psychometric tests are that they can be administered to groups in in-person or online, and they can be scored mechanically (*Ones, Dilchert, Viswesraran, & Salgado, 2017*). A limitation of most cognitive ability and intelligence tests is that they often use irrelevant content such as definition of words or symbols and multiple-choice formats. Such tests are often not seen as face valid or acceptable in organizational settings. In addition, they do not provide insights into how the individual is carrying out thought processes.

### Situational Judgment Tests and Interviews

Situational judgment measures present the examinee with a brief description of an organizational situation and ask for a statement of what action the person would intend to take (*Hough & Dilchert, 2017*). In situational judgment tests, the examinee is offered four alternative actions and asked to make a choice. In both interview and written formats, the examinee makes a statement of what he or she would intend to do but is not required to demonstrate actual behavior needed to respond to the situation. Thus, these measures are not true behavioral assessment methods. They measure procedural knowledge of what should be done but not the skills to complete a task.

### Behavioral Assessment

Behavioral assessment measures call for the examinee to use actual words or actions to respond to the situation presented. These overt behaviors show what the examinee does when faced with challenges, not just express a behavioral intention. Existing behavioral assessments include features of AWD, but they do not provide the full array of assessment information provided by a pure assessment of reasoning and interpersonal competencies offered from the proposed method.

The widely used written case study method (*Pigors & Pigors, 1961*) is often referred to as the Harvard case method used extensively in MBA programs. While there is some concern about the relevance of the case method in business schools (*Bridgman, Cummings, & McLaughlin, 2016*), the method is quite common in business schools in Iran and other countries. It asks participants to read a complex set of information and data related to the economy, local business environment, and demographics, and then to propose management solutions. Follow-up questions may give assessors

insight into reasons for proposed actions. Business case studies assess decision-making capabilities but also technical declarative information about business practices. Candidates who have specific technical information and skills (for example, financial analysis techniques or marketing practices) related to the case may have an advantage over others. By contrast AWD presents nontechnical, universal challenges to which all persons can relate. Thus, it provides a more pure measure of reasoning abilities.

By a similar analysis, other existing organizational simulations provide opportunities to assess oral communication skills but which may be contaminated with other competences. The leaderless group discussion (Bass, 1950, 1954) requires skills to interact in a small, possibly competitive group that may be dominated by a strong participant. The oral presentation exercise (Thornton & Mueller-Hanson, 2004), role play/interview simulation exercises (Fishbein & Ajzen, 1975), and in-basket/boxes (Hemphill, Griffiths, & Frederiksen, 1962) require the participant to understand organizational and interpersonal situations and exercise interpersonal skills to deal with others. The fact-finding method predominantly measures the participant's ability to stand up under challenges by the resource person (Pigors & Pigors, 1961). Additional recent information on each of these behavioral measures can be found in Thornton and Byham (1982), Thornton (1992), Thornton and Rupp (2006), Thornton, Rupp, and Hoffman (2015), Thornton and Mueller-Hanson (2004), and Thornton, Mueller-Hanson, and Rupp (2017). In summary, in contrast to these other behavioral assessment techniques, AWD provides a more pure measure of reasoning abilities.

Based on this review of the strengths and limitations of past efforts to assess behavioral competencies related to expert talent we developed AWD. We expected AWD to provide a single measure consisting of multiple dimensions of talent, to assess dimensions related to similar dimensions measured by other techniques, and to be fair to persons of different gender and age.

### Analytical Writing and Discussion (AWD)

The focus of this study is the AWD behavioral assessment exercise. It has two parts: Analytical Writing and Oral Discussion. The first part, Analytical Writing, in turn has two sections: Writing1 Analyzing an Issue calls for the candidate to read a brief passage that states or implies an issue of general interest and then to compose an essay (Appendix A), and Writing2 Analyzing an Argument calls for the candidate to read a brief passage and write a response discussing the assumptions made by the author (Appendix B). The second part, Oral Discussion, involves the assessor asking questions and making challenges about the written responses (Appendix C).

Conceptually, the AWD assesses six constructs, some

of which are assessed on other methods: It is related to the Graduate Record Examination (*Educational Testing Service, 2014*) in assessing *analytical skills, writing skills, and observation of open thinking*. The Critical Thinking Test (Facione, 1990) measures only analytical skills. Fact Finding (Thornton & Mueller-Hanson, 2004; Thornton, Mueller-Hanson, & Rupp, 2017) presents a short description of a problem in an organization to the participant, who can then ask a resource person for additional information and then make a recommendation. The resource person then challenges the participant and asks for support and rationale. Fact Finding assesses *analytical skills, active listening, oral expression, and opportunity for the assessee to learn from participation*.

### Process of AWD

The candidate is given 30 minutes to complete each written part of Analytical Writing. This time was based on pilot work; it provides enough time of full responses but puts some pressure to work efficiently, as demanded on the job. The administrator makes two copies of the written responses and gives a copy to one of two assessors. That assessor marks the two writing samples and records ratings on a 3-point scale (0, 1, 2) on each of five elements (structure, organizing, objectivity, words relevance, and grammar, each specified with behavior examples; Exhibit D1). This scale was used to provide three levels of evaluation and thus a total score 0 to 10, a range considered adequate to differentiate writing ability and that matches the scale for other ratings. These ratings are recorded separately from the Oral Discussion with the candidate and thus provide an uncontaminated assessment of writing ability. They provide evidence such as flaws or fallacies in the candidate's argument and challenges for the candidate in the next part, the Oral Discussion.

During the one-hour discussion, the second assessor meets with the candidate. The candidate introduces him/herself and reads the responses. The assessor takes notes and then asks questions and makes challenges (See Exhibit C). The assessor then rates the candidate on the eight dimensions shown in Table 1 on a 10-point scale (See Appendix D). The ratings on the two written exercises by Assessor 1, and the ratings of the oral discussion by Assessor 2 provide the bases for the evaluation of individual parts of AWD.

After the two assessors record their independent ratings on the three parts of AWD, they discuss observations and ratings and come to consensus and record the final ratings on the nine final dimensions shown in Table 6. The assessment of writing ability comes in two places, first in the independent ratings by the first assessor of the written exercises and then the consensus final rating after discussion between the two assessors.

Basing the final ratings on a consensus after this process has several advantages. The candidate has enough time

to think for 1 hour in the writing portion and then another hour during the discussion. They may learn in the challenges by the assessor, change their argument during the discussion, and make a new analysis in company with the assessor. Such actions show signs of active listening and collaboration. Having opportunities to write and discuss allow each candidate to display his/her strongest method of analysis and communication. In the dialogue, candidates are given feedback about their argument and can make revisions; assessors can see modes of thinking of the candidate.

### Assessors for AWD and Interviewers

The assessors and interviewers were people in the organization who were interested in and able to argue and discuss with others. These assessors had competencies such as critical thinking, active listening, and flexibility to learn from participants as they execute the assessment methods. The assessors are trained to conduct assessments in a consistent and standardized manner and to score the responses with reliability.

## METHOD

### Sample

The participants in this study were staff members of Mobarakeh Steel Company, Isfahan, Iran. They were being screened for selection into positions in finance, human resources, sales and marketing, production, laboratories, and so on. The common core of requirements for these positions was a set of expert talent dimensions including abilities and traits, analysis skills, and social skills. The screening process included the following stages: 12,415 people applied for announced openings; 9,772 people met age and education requirements. These candidates attended the written exam, which was based on each candidate's education (for example, industrial engineers took a 4-option test including statistics, operation research, inventory control, and planning). This yielded 219 people (197 men, 19 women, 3 no gender recorded; mean age of total = 28.66, men = 28.69, women = 28.28,  $p > .05$ ) who completed an assessment center process that included a quantitative test, interview, and behavioral exercises. The percent of the sample who are women is small but representative of the applicant pool and workforce in this organization. Eighty-four candidates were selected to continue to the final recruiting process. Final selection (yes/no) could have been used as a proxy criterion, it is not adequately independent of the other data to provide a unique measure.

### Measures

In addition to AWD, candidates participated in the following assessment methods. The evaluations of all these measures were conducted independently from each other. Interviewers and assessors did not have knowledge of

scores on any other measure, including the AWD.

The Quantitative Reasoning Measure is a multiple choice test covering basic mathematical skills, understanding of elementary mathematical concepts, ability to reason quantitatively, and to model and solve problems with quantitative methods. The Computer Project assesses computer skills by asking participants to complete a set of tasks involving Excel, Word, and PowerPoint practices. Chart Analysis simulated data analysis tasks. Candidates are given a table of data (for example, a list of 10 years of information on manpower hours for permanent employees and contractors, total hours, and production tons), asked to enter data into Excel computer, conduct analyses, and prepare one or more charts or graphs. The candidates must make the charts and prepare displays in the same room as the assessor, who asks the candidate to explain key relationships, make sensitivity analyses, and answer challenges to the results. In the Game, each group of candidates participated collectively in a simulation of tasks related to their specialty. As examples, mechanical candidates made a structure, sales and marketing candidates made a stand for an exhibition. Assessors observed candidates contributing to the activities and collaborating with others to get results. Assessors recorded evidence of how participants contribute, collaborate, pioneer, and organize activities; persuade others; and build the team to do collective tasks. The game allowed assessors to observe and rate social and problem solving skills. The Interview covered background information such as education and work experience. The interviews were competency-based and assessed behaviors of participants in their work setting. Among the dimensions assessed in Chart, Game, and Interview were selective dimensions that matched some dimensions in AWD. Table 4 shows the matches where relevant correlations appear in appropriate cells.

## RESULTS

We present two sets of results. The first set, Tables 1–4, presents analyses of initial independent ratings in the each of the three separate measures comprising AWD. The second set of results, Tables 5–7, presents analyses of the final ratings of the AWD after integration discussions between the assessors. These results show intercorrelations among the dimensions and relationships with demographic variables and other measures.

### Analyses of Three Separate Measures

Evidence of construct and discriminant validity is provided by examination of the relative size of correlations among multiple methods measuring multiple variables. In this study the methods are Oral Discussion and Writing1 and Writing2. The variables are the eight dimensions in Oral Discussion and the five elements of both Writing1 and

Writing2. It will be noted that the two writing exercises were scored by the same assessor and thus not fully independent.

Inspection of the correlations among the eight dimensions of Oral Discussion in Table 1 shows relatively large relationships, that is, over two-thirds are greater than .50, suggesting that the dimensions are measuring much the same thing. Factor analysis of these correlations was conducted using the principal component extraction method (Stevens, 2002). Convergent validity was demonstrated by the result showing that 62% of the extracted variance exceeded the critical amount of 50% suggested by Stevens. In a same way, correlations among the elements within Writ-

ing1 and Writing2 shown in Table 2 are relatively large, especially for Writing1. More importantly, the Stevens' method shows that the variance explained is 55% for Writing1 and 49% for Writing2. The results suggest that each of the three methods in AWD is measuring a single construct.

Discriminant validity is studied by examining correlations of two sets of measures of the same and different characteristics. Comparisons among these two sets of correlations are probative. Table 3 shows the correlations among measures of five elements of Writing1 and Writing2. It should reveal relatively large correlations between two ratings assessing the same element of writing ability (see values in bold print), but relatively small correlations

**TABLE 1.**

Convergent Validity of Dimensions Within Oral Discussion for Initial Ratings

	1	2	3	4	5	6	7	8
1. Flexibility		.42	.37	.43	.38	.70	.23	.23
2. Idea generation			.69	.78	.70	.52	.46	.61
3. Simplification & identifying key events				.78	.79	.53	.55	.68
4. Cause and effect analysis					.73	.53	.55	.68
5. Inductive reasoning and conclusion						.53	.50	.65
6. Active listening							.46	.41
7. Oral expression								.56
8. Collaborative problem solving								

Note.  $r > .30, p < .05$

**TABLE 2.**

Convergent Validity Within Writing1 and Within Writing2 for Initial Ratings

	Writing1			
	Organizing1	Objectivity1	Words1	Grammar1
Structure1	.28	.61	.51	.65
Organizing1		.22	.19	.27
Objectivity1			.63	.45
Words1				.35
Grammar1				
	Writing2			
	Organizing2	Objectivity2	Words2	Grammar2
Structure2	.63	.25	.35	.33
Organizing2		.13	.47	.46
Objectivity2			.20	.13
Words2				.51
Grammar2				

Note.  $r > .30, p < .05$

among ratings of different methods measuring a different element of writing. The average of correlations on the main diagonal is .54, which is comparatively larger than .29, the average of all off-diagonal values. In addition, each diagonal correlation is larger than correlations in the same column and row. This inspection supports discriminant validity among the elements within the writing exercises.

A more precise statistical indication of discriminant validity was proposed by Hensler, Christian, and Sarste (2015). Table 4 presents the correlations of the dimensions assessed by Oral Discussion with the elements measured by Writing1 and Writing2. All of these correlations are different variables measured by different methods. Inspection shows

they are quite small. Logically, they should be smaller in comparison with correlations where methods are held constant (Tables 1 & 2) or the variables are held constant (Table 3). Hensler et al. advise that a heterotrait-monotrait ratio of less than .90 indicates discriminant validity. This criterion is met for Oral Discussion in comparison with both Writing1 (.32) and Writing2 (.29). By contrast, the ratio of .90 between the two writing measures suggests they do not have discriminant validity.

The next analyses report the means and standard deviations for men and women on the ratings of all dimensions and elements of the AWD (Table 5). The means are in the mid range on the 10-point scale, and the standard deviations

**TABLE 3.**  
Discriminant Validity of Writing1 and Writing2 for Initial Ratings

Writing1	Writing2				
	Structure2	Organizing2	Objectivity2	Words2	Grammar2
Structure1	<b>.57</b>	.50	.10	.30	.37
Organizing1	.50	<b>.58</b>	.06	.17	.17
Objectivity1	.12	.19	<b>.27</b>	.17	.17
Words1	.35	.48	.12	<b>.70</b>	.52
Grammar1	.19	.39	.07	.48	<b>.59</b>

Note. **Bolded** values along the diagonal show convergent validities.  $r > .30, p < .05$

**TABLE 4.**  
Discriminant Validity Between Oral Discussion and Writing1 and Writing2 for Initial Ratings

	Oral Discussion							
	Flexibility	Idea generation	Simplification	Cause and effect analysis	Inductive reasoning	Active listening	Oral expression	Collaborative problem solving
Writing1								
Structure1	.02	.16	.23	.19	.15	.17	.15	.31
Organizing1	.06	.06	.07	.08	.11	.03	.00	.09
Objectivity1	.01	.01	.16	.14	.13	.13	.16	.25
Words1	.05	.04	.10	.04	.06	.09	.11	.14
Grammar1	.08	.28	.32	.32	.33	.23	.24	.36
Writing2								
Structure1	.10	.15	.16	.18	.16	.05	.15	.22
Organizing1	.06	.15	.24	.20	.14	.15	.15	.27
Objectivity1	.02	.07	.02	.12	.08	-.02	.03	.05
Words1	.04	.06	.11	.07	.02	.04	.06	.13
Grammar2	.17	.14	.15	.17	.02	.20	.16	.18

Note.  $r > .30, p < .05$

are sizable ( $> 1.0$ ). Significant gender differences are shown ( $p < .05$ ) on only three comparisons: Women scored higher on Oral Expression in Oral Discussion, Words in Writing1, and Words in Writing2.

Correlations of the initial ratings on all elements of Oral Discussion and Writing1 and Writing2 in relation to all ratings of Chart, Interview, and Game are available from the second author. This matrix contains a mass of information not relevant to the current issues because only a few comparisons are probative. Of more relevance are the analyses of the final ratings described in the next section.

### Analyses of Final Ratings of AWD

Table 6 includes the means and standard deviations for the final ratings on the nine dimensions comprising the AWD. These findings are important because they examine ratings used to make decisions about candidates. Note that a dimension of writing is now in this list. It is a result

of assessors discussing the two writing samples and other information from the oral discussion. The means are in the mid range of the 10-point scale and suggest the exercise is moderately difficult for the participants. Thus, the exercise permits an assessment of a distribution of competencies. This is supported by the relatively large standard deviations. Thus, there is no artificial restriction of range in scores. The correlations among the dimensions are relatively large. All correlations are statistically significant, typically in .40s and .50s, and even as large as .70, as expected from the analyses of independent initial ratings.

Table 7 shows results comparing final ratings in AWD on two demographic variables. A gender difference ( $p < .05$ ) appeared on only one dimension: Women scored significantly higher than men on Oral Expression. Although not statistically different, men scored slightly higher than women on some dimensions (for example, flexibility and idea generation), but the reverse is true for other dimen-

**TABLE 5.**

Means and Standard Deviations of Dimensions of AWD for Men and Women for Initial Ratings

	Men <i>M (SD)</i> <i>n</i> = 197	Women <i>M (SD)</i> <i>n</i> = 19
Flexible	5.57 (1.64)	5.16 (1.67)
Idea generation	4.15 (1.51)	4.05 (1.68)
Simplification & identifying key elements	4.95 (1.71)	5.16 (1.50)
Cause and effect analysis	4.97 (1.67)	5.21 (1.44)
Inductive reasoning and conclusion	4.77 (1.69)	4.39 (1.43)
Active listening	5.58 (1.49)	5.87 (1.37)
Oral expression	5.29* (1.60)	6.11* (0.89)
Collaborative problem solving	4.86 (1.55)	5.24 (1.25)
<b>Writing1</b>		
Structure1	4.68 (1.71)	5.11 (1.82)
Organizing1	5.40 (1.58)	6.05 (1.93)
Objectivity1	3.57 (2.74)	2.79 (2.30)
Words1	5.92* (1.54)	6.79* (1.47)
Grammar1	4.37 (1.87)	4.68 (2.24)
<b>Writing2</b>		
Structure2	4.97 (1.69)	5.53 (1.74)
Organizing2	5.66 (1.59)	6.26 (1.79)
Objectivity2	3.18 (2.29)	2.37 (1.89)
Words2	5.94* (1.44)	6.79* (1.62)
Grammar2	4.57 (1.85)	5.21 (1.87)

Note. \*  $p < .05$

sions (for example, simplification and cause effect analysis). These results must be interpreted with caution due to the proportionately small number of women in the sample. Examination of age effects showed only one dimension, oral expression, with a significant correlation,  $r = .20$ , with older candidates scoring slightly higher. Thus, for the most part, AWD does not discriminate against men or women or younger or older participants.

Table 7 also shows the selected dimensions of the AWD expected to be measured also by another method. For these comparisons, a correlation coefficient is reported in the relevant columns for Chart, Game, and Interview. In

all instances, the predicted relationship was significant ( $p < .05$ ), except for Active Listening measured by AWD and the Game. Thus, AWD shows convergent validity with other measures purported to assess similar dimensions.

## DISCUSSION

The AWD is a versatile behavioral method to assess an important set of competencies needed by staff experts to carry out advisory tasks when assisting line managers in organizations. Analyses show evidence of convergent validity within each of the three parts of AWD, namely, Writing I

**TABLE 6.**

Means, Standard Deviations, and Correlations Among Dimensions of AWD for Final Ratings

	<i>M (SD)</i>	1	2	3	4	5	6	7	8	9
1. Flexible	5.54 (1.63)		.42	.36	.43	.37	.70	.21	.20	.22
2. Idea generation	4.16 (1.52)			.69	.78	.69	.53	.46	.58	.42
3. Simplification & identifying key events	5.01 (1.68)				.78	.76	.53	.54	.67	.52
4. Cause and effect analysis	5.02 (1.65)					.72	.55	.54	.65	.47
5. Inductive reasoning and conclusion	4.75 (1.63)						.53	.49	.65	.53
6. Active listening	5.62 (1.46)							.46	.40	.42
7. Oral expression	5.38 (1.53)								.54	.49
8. Collaborative problem solving	4.94 (1.51)									.46
9. Writing	5.48 (1.70)									

Note. All correlations  $p < .05$

**TABLE 7.**

Means and Standard Deviations of Final Ratings of Dimensions of AWD for Men and Women, and Correlations with Other Measures

	Men <i>M (SD)</i> <i>N</i> = 197	Women <i>M (SD)</i> <i>N</i> = 19	Correlations with other measures <sup>a</sup>			
			Age	Chart	Game	Interview
Flexible	5.58 (1.62)	5.16 (1.67)	-.03	.24 <sup>+</sup>	.29 <sup>+</sup>	-
Idea generation	4.18 (1.51)	4.05 (1.68)	-.12	-	.25 <sup>+</sup>	-
Simplification & identifying key elements	4.99 (1.70)	5.16 (1.50)	-.07	.33 <sup>+</sup>	-	-
Cause and effect analysis	5.00 (1.67)	5.21 (1.45)	-.12	.31 <sup>+</sup>	-	-
Inductive reasoning and conclusion	4.79 (1.65)	4.39 (1.44)	-.01	.30 <sup>+</sup>	-	-
Active listening	5.59 (1.47)	5.87 (1.37)	.01	.20 <sup>+</sup>	.17	-
Oral expression	5.31* (1.56)	6.11* (.89)	.20 <sup>+</sup>	.33 <sup>+</sup>	-	-
Collaborative problem solving	4.91 (1.54)	5.24 (1.25)	-.05	-	.32 <sup>+</sup>	.34 <sup>+</sup>
Writing	5.50 (1.72)	5.18 (1.57)	-.03	-	-	-

<sup>a</sup> The empty cells with no correlations (-) do not provide measure of two measures of comparable dimensions in AWD and another measure. \*  $p < .05$ , <sup>+</sup>  $p < .05$

Analyzing an Issue, Writing2 Analyzing an Argument, and Oral Discussion. Convergent validity was also demonstrated between the two written exercises. Discriminant validity was demonstrated between the Oral Discussion and each of the written exercises but not between the two written exercises. Results showed the final ratings on AWD were related in expected ways to external evidence from other measures of related competencies. Furthermore, the AWD appears to be fair to demographic groups, as it is not related to gender or age.

There are several strengths of study. The participants included a large set of actual candidates for meaningful jobs in a real organization in Iran. Relatively little empirical research in human resource management has been reported in the Western literature on studies in countries in this region. The assessments were high stakes for participants, and thus we can assume they were highly motivated to perform well. This setting is in contrast with some test evaluation and validation efforts using students or even current employees. Thus, in a real world organizational setting, empirical evidence of internal and external validity was demonstrated. In addition, a variety of other measures were used to study external validity, that is, demographics and three other assessment methods including interview and two performance measures.

As with any study there were limitations. The data were collected only on a group of candidates who had been screened by other assessment methods. Although this was prudent in the practice of selecting new staff members, the range of abilities may have been restricted and thus reduced the relationships some measures may have had with others, such as age. This suggests a need to examine the function of the AWD in other settings, specifically applicants. As noted, the sample of women is relatively small. But, it must be recognized that the study was done in a manufacturing organization in Iran, a largely Muslim country with historical restrictions on employment opportunities for women, especially in certain occupations. Certainly more research should be done with larger samples of women. There was no criterion measure of job performance for this sample. Thus, although a variety of convergent and discriminant validity was demonstrated, the study was not able to provide evidence of predictive criterion validity. It would be especially informative to investigate any unique validity AWD may show in relation to other existing measures of specific competencies, for example, a purer measure of general cognitive ability as measured by some standard intelligence test. Finally, AWD is relatively labor intensive, that is, it takes more time for candidates and assessors, in comparison with other measure calling for multiple choice responses on paper or online. At the same time, a meaningful comparison will be the payoff and benefit of behavioral data from the AWD, as has been demonstrated by studies of other be-

havioral methods such as assessment centers, which have shown evidence of utility, that is, economic payoffs beyond cost (Thornton, Rupp, & Hoffman, 2015).

### Contributions and Implications

The study provides a clear articulation of competencies needed for expert talent in staff positions that do not have the formal authority of line managers. These competencies include critical thinking, reasoning, and interpersonal communication skills. The report provides empirical evidence of internal and external validity of the AWD that:

- Provides a systematic procedure to conduct a standardized and meaningful dialogue about reasoning tasks between an assessor and each individual candidate, something that is not provided by unstandardized employment interviews and certainly not by tests and questionnaires. Using writing and oral discussion together is useful because the candidates have enough time to think in the writing sections (an hour) and then they have enough time to discuss with an assessor (an hour) in the oral section. The candidates may learn in challenges made by assessor, change their argument during discussion, and make a new analysis in company with the assessor. These are signs of active listening and collaboration, very relevant to desirable actions in modern organizational settings. The method allows candidates to show abilities in both writing (one-way) and oral arguments (dialogue). In a dialogue situation the candidates are given feedback so that they can correct their arguments in real time.
- Provides a model of how other related behavioral assessment methods could be constructed. Whereas the content of Writing1 (raising a child) is quite general and the content of Writing2 (tariff) is specific to this organization, different content could be injected for other settings. Thus, the model of AWD is applicable for a wide variety of jobs, functional areas, including managers and non-managers.
- Can be used alone or in combination with other assessments of cognitive and interpersonal competencies in a program such as the assessment center method.
- Provides behavioral feedback for development of people who want to learn how to persuade others in negotiations or how to debate in business and public forums.
- Can be used to assess candidates applying to study philosophy or sociology in universities. These students need to have socio-analytical skills to persuade others.
- More broadly, it can be used to assess and develop critical thinking skills needed to debate and persuade others dealing with problems in organizations and nations.

## REFERENCES

- Bass, B. M. (1950). The leaderless group discussion. *Personnel Psychology*, 3, 17-32.
- Bass, B. M. (1954). The leaderless group discussion. *Psychological Bulletin*, 51, 465-492.
- Bridgman, T., Cummings, S. & McLaughlin, C. (2016). Restating the case: How revisiting the development of the case method can help us think differently about the future of the business school. *Academy of Management Learning and Education*, 15, 724- 741.
- Church, A. H., & Silzer, R. (2014). Going behind the corporate curtain with a Blueprint for Leadership Potential: An integrated framework for identifying high-potential talent. *People & Strategy*, 36, 51-58.
- Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *Leadership Quarterly*, 25, 63-82.
- Educational Testing Service (2014), *Official GRE® Quantitative Reasoning Practice Questions Vol 1*. Princeton, NJ: McGraw Hill Education.
- Facione, P. A. (1990). *California Critical Thinking Skills Test: College Level*, Millbra, CA: California Academic Press.
- Fishbein, M., & Azjen, I. (1975). *Belief, intention, and behavior. An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Hemphill, J. K., Griffiths, D. E., & Fredricksen, N. (1962). *Administrative performance and personality: A study of the principal in a simulated elementary school*. New York, NY: Teachers College Bureau of Publications, Columbia.
- Hensler, J., Christian, M., & Sarste, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of Marketing Science*, 43, 115-135.
- Hough, L. & Dischert, S. (2017). Personality: Its measurement and validity for employee selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp 298-325). New York, NY: Routledge.
- Kaplan Test Prep (2019). *GRE Practice Sets: 220+ Practice Questions*. New York, NY: Kaplan Publishing.
- Ones, D.S., Dilchert, S., Viswesvaran, C., & Salgado, J.F. (2017). Cognitive ability: Measurement, and validity for employee selection. In J. L. Farr, & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 251-276). New York, NY: Routledge.
- Paschen, M. & Dihmsmaier, E. (2014). *The psychology of human leadership: How to develop charisma and authority*, Berlin, Germany: SpringerVerlag.
- Pigors, P. & Pigors, F. (1961). *Case method in human relations*. New York, NY: McGraw-Hill.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin*, 124, 262-274.
- Stevens, J. P. (2002). *Applied multivariate statistics for social sciences* (4th ed.). Mahwah, NJ: Erlbaum.
- Thornton, G. C., III. (1992). *Assessment centers in human resource management*. Reading, MA: Addison-Wesley.
- Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York, NY: Academic Press.
- Thornton, G. C. III, & Mueller-Hanson, R. (2004). *Developing simulation exercises: A how-to guide for practitioners and students*. Mahwah, NJ: Erlbaum.
- Thornton, G. C., III, & Reynolds, D. H. (in press). The psychology of leadership talent potential. In T. Abraiz (Ed.), *The Routledge companion to talent management*. New York: NY: Routledge.
- Thornton, G. C., III, & Rupp, D. R. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Erlbaum.
- Thornton, G. C., III, Mueller-Hanson, R. & Rupp, D. E. (2017). *Developing simulation exercises: A guide for practitioners, students, and researchers* (2nd ed.). New York, NY: Routledge.
- Thornton, G. C., III, Rupp, D. E., & Hoffman, B. J. (2015). *Assessment center perspectives for talent management* (2nd ed.). New York, NY: Routledge.
- Weiner-Green, S. & Wolf, I. K. (2017). *Barron's GRE* (20th Ed.), Hauppauge, NY: Baron's Educational Series.

RECEIVED 05/14/19 ACCEPTED 01/19/20

## Appendix A

## Writing1: Analyzing an Issue

You will be given a brief passage that states or implies an issue of general interest. Instructions will be given on how to respond to the issue. Your response will be graded by how well you:

- respond to the specific task instructions,
- consider the complexities of the issue,
- organize, develop, and express your ideas,
- support your position with relevant reasoning and examples,
- control the elements of standard written expression.

"It takes a village to raise a child." The education of your children is the task of the community as a whole not merely the province of teachers and local school administrators.

In 30 minutes, compose an essay on the quotation above. Discuss the extent to which you agree or disagree with the views expressed in the quotation. In doing so, be sure to explain the reasoning behind your taking this position. Consider to what degree the statement may or may not be true; take this into account in your discussion. You may not write on any other topic.

(Weiner-Green & Wolf, 2017)

## Appendix B

## Writing2: Analyzing an Argument

You will be given a brief passage that presents an argument. Instructions will be given on how to respond to the argument. Your response will be graded by how well you:

- respond to the specific task instructions,
- identify and analyze important features of the passage,
- organize, develop, and express your analysis,
- support your position with relevant reasoning and examples,
- control the elements of standard written expression.

Respond to the instructions below and support your position with relevant reasoning drawn from your academic studies, reading, observation, and/or experience.

Argument: One increasingly popular policy for promoting renewable energy is a feed-in tariff. Under such a policy, investors on any scale, from large corporations to individual homeowners, produce their own energy from solar panels installed on their property. Electricity companies are then required to purchase the energy through a long-term contract at an increased rate that would allow the investors to more than offset the cost over time. There is no denying that the initial cost of solar installation would be a burden on the investor. In strenuous economic times, both businesses and homeowners might be reluctant to make the investment, with concern that the payout could be less than sufficient or the plan might prove unfeasible. However, research has shown that a feed-in tariff plan is not only stable but also exceptionally effective, and ought to be more actively pursued.

Write a response in which you consider the assumptions made by the author in the passage. Discuss how these assumptions affect the validity of the argument and the possible implications should these assumptions be proved wrong.

(Kaplan Test Prep, 2019)

## Appendix C

## Questions and Challenges for Oral Discussion

1. What is your opinion about this conclusion?
2. How did you conclude?
3. What are the assumptions?
4. What are the implications?
5. Please give me more reasons?
6. Why don't you think that the other sides are true?
7. What are the key points?
8. If I think your reasoning is wrong, how would you persuade me?
9. After listening to my questions and reviewing my reasons, please read your analytical writing again and tell me if your conclusion changed or not.
10. Don't you think that your analytical writing need to be corrected?
11. Please summarize our discussion.
12. What did you learn in our discussion?

## Appendix D

## Rating Scale for Five Elements of Analytical Writing

	Feature	Description	Rating
1	Structure	The writing has introduction, body, and conclusion.	0-2
2	Organizing	The writing gives a big picture, main ideas and supports; categorizing and ordering them.	0-2
3	Objectivity	The writing gives examples and objective reasoning.	0-2
4	Words	The writing contains correct, relevant and meaningful words.	0-2
5	Grammar	The writing has correct formation of sentences, tenses, and conjunctions.	0-2

## Rating Scale for Dimensions of AWD

- 9, 10 Excellent
- 7, 8 Strength
- 5, 6 Acceptable
- 3, 4 Weak but improvable
- 1, 2 Not improvable