

CHAPTER 6

TESTING FOR PROMOTIONS

FINNING LTD. AND IAMAW LODGE 99 CHARGEHAND SELECTION GRIEVANCE

I. MANAGEMENT PERSPECTIVE

WILLIAM J. ARMSTRONG*

Background

Finning Ltd. is a heavy equipment dealer based in Vancouver, British Columbia. Its main product line is Caterpillar equipment and it has branches in parts of Canada, Chile, Poland, and the United Kingdom. Finning has branches in eight Alberta cities and two Northwest Territories locations. (There are multiple branches in Edmonton and Fort McMurray.) There are approximately 800 employees in these branches, of which approximately 600 are in the bargaining unit.

The International Association of Machinists and Aerospace Workers, Local Lodge 99 (IAMAW), represents the bargaining unit employees at Finning for Alberta and the Northwest Territories. It was formerly the Independent Union of Heavy Equipment Trades, which merged some years ago with the IAMAW. Finning's British Columbia employees are represented by a different local of the IAMAW.

R. Angus Caterpillar was the Caterpillar dealer for Alberta and part of the Northwest Territories and was purchased by Finning in 1989. Since purchasing R. Angus, Finning has negotiated amendments to the collective agreement to introduce into Alberta two job classes from the British Columbia collective agreement. One was a nonticketed position in parts called "material supply assistant." The other was a supervisory position called a "chargehand." The

*Principal, Laird Armstrong, Calgary, Alberta.

chargehand position is the highest paying position in the bargaining unit, above the previously existing position of leadhand. In brief, the company saw the chargehand position as a stepping stone into management, and the union saw the chargehand position as the leadhand with a different name.

Chargehand Selection Process

When the chargehand position was added to the Alberta collective agreement, Finning decided upon a selection process that reflected its conception of the position as a stepping stone into management. The company included psychological testing as part of the selection process. The company met with the union to explain the process it intended to use.

The selection process began with a general invitation for interested employees to apply for possible chargehand positions. Interested applicants were given three psychological tests and were interviewed in what Finning calls an "interview circuit." In an interview circuit, each applicant is interviewed one-on-one by several managers in sequence. The managers later meet and rank the applicants. The tests and interviews were conducted in one day at either the Calgary or Edmonton sites. This prescreening resulted in a shortlist of individuals to fill the positions.

Chargehand positions were then posted at various Finning branches. After a union grievance, Finning agreed that these postings were open to all employees, including those who had not participated in the prescreening or who had participated but not made the shortlist. The selection process at each branch was under the discretion of the local managers. In some cases, a new interview circuit was held and conducted by local managers. In other cases, success on the prescreening largely dictated the outcome.

The union took issue with the use of psychological tests and three of the selections on the actual postings. In the first case, the grievance was sustained, and the grievor was awarded the position. In the second case, the grievance was sustained in part, and the employer was directed to reconsider. In the third case, the grievance was dismissed on the basis that there was ample evidence to support the employer's decision despite the flaws in the selection process.

Collective Agreement

Article 5.05a of the collective agreement stated:

In filling job vacancies, including promotions and new positions, the job shall be awarded within thirty (30) calendar days of the posting to the most senior applicant provided such applicant is qualified to perform the job. Qualification being defined as possessing a proficiency certificate where necessary, training, knowledge, and experience and past performance (of which the Employer shall be the sole judge). Unsuccessful applicants will be notified and given the reasons [why] they were not accepted for the posting applied for.

There was also a letter of understanding concerning the implementation of the chargehand classification. It stated, in part:

All Chargehand positions are to be filled using the appropriate interview circuit. The appropriate selection criteria will be established and explained. Selections for any positions filled will be from available Alberta bargaining unit Employees.

There were arguments on the relationship between Article 5.05a and the letter of understanding, as well as on the meaning of the phrase "of which the Employer shall be the sole judge."

Psychological Tests Used

The company utilized three tests in the selection process:

1. Wonderlic Personnel Test
2. The Common Sense Manager (TKIM)
3. Thomas International Management Systems Ltd. Personal Profile Analysis (PPA)

At the hearing, evidence about the tests was heard from three individuals. The union called an Assistant Professor of Psychology from The University of Calgary, Theresa J.B. Kline, Ph.D., who reviewed and commented on the tests. The employer called Larry Stefan, Ph.D., a chartered psychologist, who had consulted to the company on higher management selections but not on the design of the selection process. The employer also called Sean Magennis of the Canadian branch of Thomas International. Both psychologists discussed testing in general and the three tests used in this case. Both referred to the distinction between tests that may be administered only by psychologists and those that may be administered by lay persons. All three tests are of the latter type.

Company Argument on the Tests

In the arbitration, the company contended that in addition to management's right to determine the specific qualifications necessary for a particular job, arbitrators have recognized a companion management right to measure those qualifications by administering examinations and tests to job applicants. Such a right is implied in the absence of any provision to the contrary in the collective agreement. The company cited *Re Oil, Chemical & Atomic Workers Local 9-14 and Polymer Corporation Ltd.*, in which the arbitration board stated:

There is no doubt that in general, an employer may properly require employees who apply for a posted job to submit to a test to determine their ability or qualifications. . . . Employees may be required to meet appropriate standards of performance, of basic skill or ability, or of general knowledge or aptitude. Whether the test produces results which may properly be relied upon is a matter of fact to be determined in each case. It must, of course, be administered fairly and without bias, but it must as well meet certain standards of reliability and validity.¹

The company also cited *Re Domtar Inc. and U.E.W.*, wherein the arbitration board stated:

Turning to the question of the test itself, in the absence of collective agreement language to the contrary, arbitrators have consistently recognized management's right to use a test to determine or measure an employee's qualifications or ability for the position in question, subject to certain principles and factors applied to the testing process.²

Nothing in the collective agreement limited or abrogated the right of the company, recognized in the above authorities, to require employees to submit to tests to permit proper assessment of applicants' qualifications for the chargehand positions. However, as the passages cited indicate, management's right to test job applicants is not an unqualified one. Arbitrators have developed certain standards, described as follows in Brown and Beatty:

In order to assess the requisite ability and qualifications of employees for a particular job, arbitrators recognize that it is quite proper for an employer to require the applicants to submit to examinations and other tests, including general aptitude tests, to demonstrate their skill and ability. Those tests must be administered fairly, without bias, and meet certain standards of relevance, reliability and validity. Specifically, in order that such tests may be said to reasonably reflect an

¹(1968), 19 L.A.C. 386.

²(1986), 22 L.A.C. 3d 247, 252.

employee's ability and qualifications, arbitrators routinely inquire into the reason for the institution of the test, the adequacy of the preparation that was afforded to the employee prior to the test, the method and circumstances under which the test was administered, the reliability of the marking of the test, and the relevance of the test to the particular work to be performed. Accordingly, if it were determined that a test was not administered consistently, that different questions were asked of each employee, that the results had become stale, that the test or parts of it were not designed to elicit any information relevant to the job in question, that there were no fixed levels established within the test which an employee had to meet to demonstrate sufficient ability, or that the marking system was highly subjective, any decision by an employer based on such tests as to the relative abilities of the applicants would be found to be unreasonable and improper. As well, it seems generally agreed that such tests ought not to be used by the employer as the sole basis for its assessment of an employee's qualifications where there is other valid objective evidence available to it.³

It was submitted that the tests and the manner in which they were administered in this case satisfied each of the concerns set out in *Brown and Beatty*, as typically addressed by arbitrators in determining whether a particular test reasonably reflects the abilities of applicants.

1. *Reason for administration of test.* In this context, the parties had negotiated for the inclusion of a new position in the collective agreement. Both the novelty and importance of the chargehand position, on the one hand, and the nature of the required qualities, on the other, made past performance an unreliable indicator of the likely success in the chargehand position. As a result, the tests were determined to be the most appropriate method of assessing some of the qualities required for the position.
2. *The adequacy of the preparation that was afforded to the employee prior to the test.* In May 1993, the company announced forthcoming postings for the chargehand positions. The union was made aware of the selection criteria in information that described the implementation timetable for filling the available positions and that set out the three testing instruments to be used, with brief descriptions of each. It should be noted that the type of tests at issue are not tests that require or permit specific preparation on the part of employees.

³*Brown & Beatty, Canadian Labour Arbitration*, 3d ed. (Canada Law Book 1996), Sec. 6:3340.

3. *The method and circumstances under which the test was administered.* There was nothing in the method or circumstances under which the tests were administered that undermined the validity or reliability of the testing process.
4. *The reliability of the marking of the test.* There was no evidence that there was anything unreliable about the marking of the tests.
5. *The relevance of the test to the particular work to be performed.* In addressing this issue, it must be kept in mind that it is not the role of the arbitration board to hold the employer to a standard of perfection. Rather, it must be determined whether the employer acted *reasonably* in relying on the tests in question as indicators of the abilities and qualifications they were intended to measure.⁴

The Wonderlic Personnel Test

The Wonderlic is a test of general cognitive ability or what is commonly called "intelligence." This test can be administered in 12 minutes and scored by a lay person. It is therefore practically suited to the use by a company needing a short but reliable test of general cognitive ability.

Both psychologists recognized the Wonderlic as a widely used test with a well-recognized history in the psychological literature. Both agreed that the use of this test as part of a selection process involving interviews, reference checks, and other possible steps was appropriate.

Where the psychologists differed somewhat was regarding the relative weight to be given to tests of general cognitive ability in a selection process. Citing research by Dr. Hunter, Dr. Stefan indicated that tests of general cognitive ability are the best predictor of future job performance, especially for more cognitively complex jobs such as supervisory jobs. Although Dr. Kline also cited research by Dr. Hunter, she would not place as much emphasis on a test of general cognitive ability.

The TKIM

The TKIM is based on the construct that a person may have "tacit knowledge" or "street smarts." This test endeavours to measure the

⁴See Scherman, *Re R.W.D.S.U., Local 480 and Beverage Central Ltd.*, unreported (1 May, 1990).

tacit knowledge of an applicant relative to the tacit knowledge of a group of managers from the same company. Both psychologists indicated there are a large number of such tests available for administration by psychologists only and a larger number available for administration by lay persons.

Dr. Stefan testified that to the extent the TKIM measures general cognitive ability, it would be a valid predictor of future job performance.

Dr. Kline's reservations were that the construct of "tacit knowledge" is not generally accepted in the academic community and that she could find only one study of the reliability and validity of the TKIM at the time she wrote her report. Dr. Kline indicated she had found a second study of the TKIM just the day before she testified. Dr. Kline's position was that she would not recommend the use of any test that had not been extensively studied in published reports. Therefore, she had reservations about the TKIM because only two studies had been published on it.

The PPA

The PPA is a measure of personality. It can be used alone or in comparison with a standard profile for the position created by a management group in the company. This standard profile was not used by Finning.

Dr. Stefan indicated that a number of his clients used the PPA. Mr. Magennis gave evidence that the test is used in 34 countries, that it has been translated into 22 languages, and that more than 100,000 Canadians and 250,000 Britons had taken the test last year.

Dr. Kline was very critical of the PPA in her report. She termed "absurd" the conclusion in the Kaplan Report (a paper provided by Thomas International) that it was a valuable instrument. In oral evidence, she said her *initial* reaction was that it was created and marketed by "charlatans." She further testified that she had considerably revised her opinion upon review of material provided by Mr. Magennis just prior to her testimony. Upon reviewing this material, her opinion changed to essentially the same position as she took on the TKIM, that is, that she would not recommend a test not extensively studied in published reports.

The company submitted that Dr. Kline's standard was unreasonably strict. A company cannot be expected to employ a psychologist to administer tests for every position in the way it might for its most senior executives. The cost of such a practice would be prohibitive.

Once the psychologist is excluded, those tests that can be administered only by psychologists are also excluded. For example, if a company employs a psychologist, the psychologist could administer a test such as the Minnesota Multiphasic Personality Inventory (MMPI). (The MMPI was one of the tests with which the PPA was compared in the Kaplan Report.) If a psychologist is not employed and the company still wishes in some way to test personality, it cannot use the MMPI and so must use another test administrable by a lay person, such as the PPA.

In assessing the use of tests by Finning, it was important to examine the role played by tests in the selection process. First, the company interviewed all applicants in an interview circuit as specified in the letter of understanding. Then, before the tests were scored, the interviewer established a tentative division of the applicants into "yes," "maybe," and "no" groups. The effect of the test scores as well as reference checks was to move all "possibles" into the "yes" group and to remove one person, Clayton Lokos, from the "yes" list.

The tests and reference checks were clearly secondary to the interview process and were used as corroboration for the decision reached on the basis of the interview circuit. Stan Prince testified about his decision to award the chargehand position in Fort McMurray to Mark Meldrum instead of Clayton Lokos because Mr. Meldrum was on the shortlist and Mr. Lokos was not. In Mr. Prince's view, the decision was a nine on a scale of one to ten.

In the selection at the Reman Centre, tests were of even less influence, as a second interview circuit was done. All applicants were interviewed. Of the six shortlisted from the second interview circuit, three also had been shortlisted on the preselection.

The union's position was that seniority and past performance should be the only relevant factors considered, especially when promoting internally. Yet the collective evidence of the psychologists was that the predictive validity of tests is greater than that of structured interviews, which are better than unstructured interviews, which are better than reference checks (i.e., past performance), which are better than seniority.

It was submitted by the company that the circumstances before the arbitration board revealed none of the shortcomings that have been identified in the case law as undermining the testing process.

In fact, the tests operated to bring objective balance to the assessment of largely subjective qualities. The desirability of employing tests for that very reason was noted in *Re Corporation of City of Stratford and Stratford Professional Firefighters' Association, Local 534 International Association of Firefighters* as follows:

While testing may be inappropriate in some situations it is my view that there are numerous situations where testing is desirable. Tests serve to remove the subjective element from assessment and provide an objective measure of qualifications. Also, a proper testing procedure should allay any suspicion among competing employees that bias or discrimination exists in awarding jobs. In sum, tests are a valid tool which may assist an employer in exercising reasonable judgment as to the qualifications or abilities of competing employees, subject to the qualification that this should not be read as constituting a blanket approval for tests in all circumstances.⁵

Here, the tests were devised by experts, have been recognized as useful tools in evaluating the qualities for which they were used, were administered in like fashion to all employees, and were scored/interpreted correctly by experts. No one test served as the sole basis for making a decision as to who would and who would not be successful in the chargehand position. The company administered the tests, obtained references, interviewed, and, in some cases, reinterviewed the applicants for the positions. Following receipt of the test scores, several applicants were added to the shortlist. It was clear that the tests, in conjunction with the interview circuits and reference checks, provided a thorough measure of the various qualifications required of the applicants for the chargehand positions. Under these circumstances, it was submitted that there was no basis upon which the board could fault the company's use of the testing process.

Union Argument on the Tests

I will leave to Mr. Carpenter's paper the union argument on the use of tests.

The Arbitrator's Award

The arbitrator found the company's selection process to be well-intentioned, reasonable in principle, without evidence of bias or

⁵25 L.A.C. 2d 170, 177.

manipulation, but nonetheless flawed in several respects. The arbitrator noted as follows:

- The company should conduct a thorough job analysis before any further chargehand selections.
- There were numerous instances of assessments and decisions by interviewers that were inappropriate or even incomprehensible.
- The Wonderlic was the best assessment tool of the three used, but it should not be used in isolation.
- The use of the TKIM and PPA should be reconsidered. The PPA and probably the TKIM should be used to provide information for the interviews. The PPA should be used with a "Human Job Analysis," which derives from a thorough job analysis.
- A job selection process should include interviews, reference checks, evaluations as well as tests, with each resource made as valid as possible.

Comments on the Award

Although an application for judicial review was filed by the union, it has not proceeded, and it appears that it will not. I therefore feel free to comment on the award.

It seems to me that employers are often caught between the proverbial "rock and a hard place" in attempting to move away from the traditional unstructured interview in job selection. Experts (including those who testified in this case) seem to agree that the traditional interview has a very low reliability as a predictor of future job performance. Human resource managers may seek to improve the reliability of the job selection process by using psychological tests or structured interview techniques. However, in doing so, the inevitable result is a process that is both more complex and more open to challenge by a union.

My experience in both this case and other cases involving structured interviewing techniques is that the result of these attempts at improvement is a selection grievance hearing that is longer, more expensive, and often involves expert witnesses. Line managers generally will not be able to apply tests or structured interview techniques as well as experts, yet it is not economical to employ experts for all job selections.

The company in this case sought to minimize the subjectivity of the process by the use of both interview circuits and psychological

tests. There are many tests from which to choose, even after elimination of those that can be used only by a trained psychologist. In this case, it offers small comfort to the company when two of the three individual grievances are upheld in whole or in part and when two of the three tests are questioned. (The factors that resulted in the third grievance being denied were factors that did not arise out of the use of either the tests or the interviews.)

Arbitrators should recognize that job selections are made in the real world by line managers who are not experts in job selection. Greater recognition of these realities should be given where job selections are made in good faith, without bias, and in a sincere attempt to reduce the subjectivity of the process.

II. UNION PERSPECTIVE

JOHN CARPENTER*

Facts

The employer had for some time attempted to bargain a new bargaining unit position, which it called a chargehand. This was to be a position above the current leadhand position and for which the employer wanted separate seniority entitlement. A letter of understanding was eventually entered into that created the new position in 1992, but without the separate seniority entitlement.

The union and the employer had a collective agreement in effect that contained a provision on promotions that stated:

In filling job vacancies, including promotions and new positions, the job shall be awarded within thirty (30) calendar days of the posting to the most senior applicant provided such applicant is qualified to perform the job. Qualification being defined as possessing a proficiency certificate where necessary, training, knowledge, and experience and past performance (of which the Employer shall be the sole judge.) Unsuccessful applicants will be notified and given the reasons [why] they were not accepted for the posting applied for.

The employer gave notice of an interview circuit from which it would create a shortlist of eligible candidates for the new position. In July 1993, the employer conducted the interviews and administered the tests to the interested applicants. In August 1993, the

*Partner, Chivers Greckol & Kanee, Edmonton, Alberta.

employer advised all applicants whether or not they had made the shortlist. The current leadhand in the Mildred Lake Shop did not make the shortlist. The Mildred Lake Shop was one of the busiest in the company's operation. The leadhand of the Town of Fort McMurray Shop, as well, failed to make the shortlist. Two other members who had been leadhands did make the shortlist but were none the less passed over for a far less senior applicant in one of the first chargehand postings awarded after the interview circuit. When the leadhand of the Mildred Lake Shop insisted on applying for the chargehand posting in his shop despite the fact that he had not made the shortlist, the company withdrew the posting.

A number of grievances were filed covering the interview process, the testing itself, the manner in which the tests were used, and the awarding of the positions themselves. Alan W. Beattie was appointed as a single arbitrator in respect to all the grievances and heard the matter over six days of hearing in the fall and winter of 1994 and 1995. Written submissions were subsequently filed, and a decision was rendered on July 28, 1995.¹ An application for judicial review was filed by the union in respect to the decision and that matter was resolved with a letter of understanding signed in November 1995. By that letter, all interview documents and comments, as well as the results of two of the tests and the acceptance and/or rejection memo, were removed from the original applicants' files. In addition, the shortlist that had been created was abolished, and it was made clear that future chargehand positions would be open to all applicants.

Since then, the parties have attempted, without success, to develop a new promotion procedure acceptable to both. In the current round of bargaining, which has resulted in the first strike by the union in the company's history, one of the most contentious issues has been the demand by the company to amend the promotion article to eliminate threshold qualifications and the seniority protection. At the Mildred Lake Shop, no chargehand position has ever been created or filled, and the grievor denied the promotion by both the employer and arbitrator remains the leadhand.

¹*Finning Ltd. v. I.A.M., Local 99 and Dhillon and Lokos and Waye*, Alta. G.A.A. 95-012 G (July 28, 1995) (Beattie).

The Union's Submissions on Law

Seniority Rights

Seniority rights are one of the bases for the foundation of the collective bargaining system and an integral part of collective agreements. Seniority provisions are designed to grant certain preferences to employees based, in whole or in part, on their accumulated length of service. The theory underlying such systems is to provide those employees possessing the longest record of service with, in the context of promotions, the greatest potential for advancement. In reviewing a promotion procedure, an arbitrator must keep in mind the fundamental nature of seniority rights.²

As stated by Arbitrator Reville in the *Tung-Sol* decision:

Seniority is one of the most important and far-reaching benefits which the trade union movement has been able to secure for its members by virtue of the collective bargaining process. . . . It follows, therefore, that an employee's seniority should only be affected by very clear language in the collective agreement concerned and that arbitrators should construe the collective agreement with the utmost strictness wherever it is contended that an employee's seniority has been forfeited, truncated or abridged under the relevant section of the collective agreement.³

Where the parties have explicitly recognized the paramountcy of seniority in promotions procedures, an employer denial of such benefit to an employee should be reviewed thoroughly and strictly. The employer should bear a heavy onus to justify its decision in such context.

The Collective Agreement Promotion Provision

The relevant provision, Article 5.05a, is a sufficient ability clause where a senior employee is entitled to a position if he or she is able to perform it. As stated by then Arbitrator Laskin in the *Westeel Products* decision, "In such case, a senior man who is equal to the job is entitled to it, although there may be a junior applicant who can do it better."⁴

²Brown & Beatty, *Canadian Labour Arbitration*, 3d ed. (Canada Law Book 1996), 6-1.

³*Tung-Sol of Canada Ltd.* (1964), 15 L.A.C. 161, quoted in Brown & Beatty, *id.* at 6-1.

⁴*Westeel Products Ltd.* (1960), 11 L.A.C. 199, quoted in Brown & Beatty, *supra* note 2 at 6-36.

By this type of clause, the qualifications of the employee are evaluated according to requirements of the collective agreement and the legitimate requirements of the job. The proper approach was set out by Arbitrator Bowman, who stated that:

There is no question that this article introduces what is commonly referred to as a "threshold clause" for promotion or hiring. This means that where there is more than one candidate for the position the candidates are *not compared with each other*, for the purpose of determining which may be the most skilled, or most qualified, but rather, that *each is compared to an objective standard indicating the requirements for successfully carrying out the position*. Of the persons who demonstrate capacity to meet the position's requirements, with perhaps some limited training and generally some familiarization time, the most senior will be given the position over other qualified applicants. In other words, among those who can do the job, the most senior is entitled to receive it.⁵

The employer must be correct in complying with the requirements of the collective agreement. The arbitral standard of review in this portion of the inquiry, as noted by Arbitrator Hope, is one of correctness, not reasonableness.⁶

Responsibilities of Grievor and Employer

Commentators have written that in cases where a "threshold" clause is at issue, the onus is on the grievor to establish that he or she possesses the requisite qualifications for the job in question. The employer in the face of that evidence will generally come forward and establish the grounds for its opinion. The overall onus nonetheless remains on the grievor to prove that he or she can perform the job even though perhaps not as well as a less senior candidate.⁷

However, the grievor's onus to establish that he or she has the qualifications necessary to do the job is in the context of the arbitration board's review of the promotion process. In other words, the grievor's onus comes into play in the context of the employer having shown (or the board having satisfied itself) that

⁵*Re Northwest Child & Family Services Agency and C.U.P.E., Local 2153* (1990), 12 L.A.C.4th 383 (Bowman) at 384 (emphasis added).

⁶*Re British Columbia (Workers' Compensation Board) and Workers' Compensation Board Employees' Union* (1989), 4 L.A.C.4th 141 (Hope) at 167.

⁷*Id.* at 164.

the qualifications were set according to the requirements of the collective agreement, and that the procedure used by the employer was valid, fair, and appropriate. For this reason, the onus is on the employer to validate the test used. As stated by Arbitrator Weiler, "... the employer must show that his tests are a real measure of the ability actually required for the job as it will be performed."⁸

Not surprisingly, a grievor who has worked for a lengthy period of time in a position comparable to the one to which he or she seeks to be appointed, must be seen *prima facie* as having the abilities and qualifications to perform in the position.⁹

The Employer's Criteria and Procedure

It is firmly settled that an employee's claim of being improperly denied a particular job will prevail if it can be established that the standards or criteria relied upon by the employer were not contemplated in the collective agreement or that the procedure used was not valid, fair, or appropriate.¹⁰

A useful methodology to approach the assessment of the employer's decision has been set out by Arbitrator Fraser in the *Northern Telecom* case. The decision, which has been adopted in several cases, set out a three-part test as follows:

1. Was there compliance with the collective agreement? (i.e., were the relevant standards in the collective agreement, and only those standards, applied?)
2. Was the procedure by which these standards were applied fair, appropriate, and unbiased? (i.e., were the specific job requirements, and any methods of evaluation or testing, appropriate for the competition, and not subjectively biased in favour of one applicant, or discriminatory against another?)
3. Was the decision a reasonable one? (i.e., given compliance and a fair procedure, was the result, objectively viewed, one that falls within a range of reasonable decisions?)¹¹

⁸Mackenzie, *Testing at Arbitration: The Employee's Perspective* in Labour Arbitration Yearbook 1993 (Butterworths-Lancaster 1993), 198-99. See also Gorsky, Usprich & Brandt, *Evidence and Procedure in Canadian Labour Arbitration* (Carswell 1994), 9-22-9-23.

⁹*Re C.B.R.T. and St. Lawrence Seaway Authority* (1969), 23 L.A.C. 156 at 161.

¹⁰*Supra* note 6, at 167.

¹¹*Supra* note 2, at 6-56.

Employers Failure to Comply With Article 5.05a

Article 5.05a expressly defines the components of "qualification":

1. possessing a proficiency certificate where necessary,
2. training,
3. knowledge, and
4. experience and past performance.

The collective agreement directs that these factors (and only these factors) are to be considered in filling job vacancies, including promotions and new positions. The criteria used by the employer and the procedures used to determine those criteria must address these factors, and only these factors.

As stated by Arbitrator Ponak in the *Calgary District Hospital* case:

... It is an Employer's prerogative to determine the necessary requirements of a job and then to develop a selection process that fairly measures such requirements. Its actions, however, must be consistent with the collective agreement. For example, if the collective agreement stipulates that "ability" is to be a key criterion in selection, it is incumbent on the Employer to develop a measure of ability that is rooted in the attributes of the job. Thus, the job selection process involves combining the inherent characteristics of the job with the guidelines of the collective agreement. In the instant case, it is the Board's conclusion that, although the Employer was fully cognizant of the nature and demands of the Nurse Clinical position, it failed to properly relate the characteristics of the job to the requirements of the collective agreement.¹²

In this matter, the employer failed to address the requirements of Article 5.05a in fashioning the assessment procedure, and, consequently, the elements evaluated did not correspond to those in the collective agreement.

Unfair and Inappropriate Procedure

The Employer Failed to Conduct a Balanced, Complete Process. An employer's selection procedure should be tested for its completeness.¹³ All relevant factors must be assessed by the employer; irrelevant factors should not be considered. Further, the employer

¹²*Re Northern Telecom Ltd. and United Automobile Workers, Local 1839* (1980), 25 L.A.C.2d 379 at 384.

¹³*Calgary District Hospital Group (Holy Cross Hospital) and United Nurses of Alberta, Local 121* (Meehan), Alberta, Ponak, Chair, April 21, 1989 at 14 (unreported).

must use all relevant sources of information and must not rely on an interview or test to the exclusion of other sources.¹⁴

The experts who gave evidence at the arbitration hearing, despite their differences, were in agreement that a balanced approach in the selection process is advisable:

- *Union Expert:* It is widely acknowledged that any single test or battery of tests should comprise only one aspect of the entire selection system. There are many other sources from which to obtain valuable information about an applicant or employee, such as past work performance records, peer evaluations, education and training records, and past experience. All of these, and other selection/assessment tools should rightfully be part of the selection system when appropriate.
- *Company Expert:* A psychologist would still use a balanced approach; you wouldn't want to put your client at risk.¹⁵

Guidance as to what constitutes a complete and balanced selection process may be sought from the collective agreement and arbitral jurisprudence on the issue. The elements listed in Article 5.05a of the collective agreement suggest that a complete and balanced process would include a review of proficiency certificates, education and courses, experience in similar jobs, history with the employer, annual or other regular appraisals, review of familiarity with employer equipment, and feedback from current or prior supervisors.

The evidence showed that the candidates participated in an interview, from which a shortlist was generated. The interview was clearly a major threshold. The interview questions, however, tended to address only soft management issues such as management style, setting goals, and dealing with problem employees.

The interview process excluded readily available, objective sources of information on the applicants. It is clear that where an interview is the predominant determinant in a selection process, the validity of the selection process may be called into question.¹⁶ The learned

¹⁴*Supra* note 2, at 6-36.3.

¹⁵Union expert material from the report "Re: Review of Testing Procedure and Evidence at Grievance Arbitration" by Dr. Theresa Kline. Company expert material from notes of company experts testimony.

¹⁶Mackenzie, *supra* note 8, at 195-96.

Arbitrator Cherniack provided an extensive review in the *Fairview Home* decision as follows:

An interview can be an artificial assessment of an applicant's ability to talk, to charm, or to use words that the interviewer clearly wants to hear. The ability to articulate, or the state of being excited about the prospect of becoming a charge nurse, does not necessarily prove an ability to *be* a good charge nurse.

* * *

Interviews cannot and should not be used, however, as a complete method of assessment. The ability to conduct oneself during an interview is only one facet of an employee's abilities, and often it is not a particularly significant or relevant facet. So much depends, unfortunately, on the ability of the interviewer to go beyond the surface impressions in the artificial atmosphere of an interview, and probe deeply into the applicant's vision and knowledge. . . .

The Employer did not use the interview in that creative way. In effect the interview became a test, without any notice being given to the applicants that they would be tested and on what aspects they would be tested. There is a wealth of arbitral jurisprudence on testing. Suffice it to say that a test, if used, must be reasonably related to the job in question, and that the use of a test to the exclusion of all other factors can be grounds for overturning the decision of management. In the words of arbitrator Freedman (*Re Winnipeg (City) and C.U.P.E., Loc. 500* (1990), 12 L.A.C. (4th) 231 at pp. 242-3. Mr. Freedman goes on to cite with approval arbitrator Hope in *British Columbia (Workers' Compensation Board)*, *supra*, on the issue of relying "almost exclusively on the interview process":

A balanced assessment is required and all relevant objective evidence must be taken into account. Undoubtedly, the failure of these two grievors to pass the tests is a significant consideration, but it is not the only consideration.

.

To impose as a mandatory requirement that the test be passed, and to decide that if it is not passed, all other factors, however positive, including performance as an acting foreman, would effectively not be taken into account, is an inappropriate method of evaluation.

The employer used the interview as a sole method of comparing the two employees, and in doing so it acted unfairly and unreasonably.¹⁷

Arbitrator Freedman in the *Winnipeg City* decision provided further pertinent observations as follows:

¹⁷*Re Elgin (Country) Roman Catholic School Board and London & District Service Workers' Union, Local 220* (1992), 26 L.A.C.4th 204 (Ont. Rose) at 208.

In my judgment the Agreement requires that the City consider all relevant indicators of ability. No doubt test results, especially from a test as carefully put together as this one, are very significant indicators of ability. But there are other indicators and when the City effectively disregards those other indicators because of a failure to achieve the pass mark on the test, the City has not properly evaluated the applicants. It is, for example, possible that very positive indicators derived from the interview, past performance and references could outweigh the fact that an applicant almost achieved a pass mark, but did not quite do so. That is a judgment call that the City ought to make, but when it established passing the test as a *sine qua non* the City deprived itself of the opportunity to consider the other factors in an appropriately balanced way.¹⁸

The employer in this instance did not make proper investigation of or take into account the grievors' past experience, did not utilize performance appraisals, and performed only a limited reference check. Failure of the employer to consider experience and past performance for the candidates is not only a failure to comply with Article 5.05a of the collective agreement, but also a failure to conduct a complete assessment.

The Tests Used by the Employer Were Not Valid or Reliable. Tests must meet certain standards of relevance, reliability, and validity.¹⁹ A useful overview is provided by Professor Rigg as follows:

Complaints about the particular test usually relate to the validity of the test, *i.e.*, the ability of the test to measure the required abilities or aptitudes, and to the reliability of the test, *i.e.*, the extent to which different parts of the test measure the same aptitude and the extent to which the test is consistent over time. Arbitrators have held that a selected test must measure an aptitude required for the position and have allowed the grievance where it did not. In determining whether a test is valid, arbitrators may hear evidence from industrial psychologists or other experts in employment testing, or they may be referred to journal articles or books which describe the test in question. Additionally, where the test has been administered to individuals, including the grievor, who have satisfactorily performed the job, the arbitrator may look at how those individuals performed to decide whether the test bears a sufficient relationship to the position. Of course, where the position is a new one in the employer's operation, that information may not be available and the arbitrator must fall back on expert testimony as to whether the test adequately relates to the position.²⁰

¹⁸*Re Fairview Home, Inc. and Fairview Nurses N.M.U., Local 21* (1991), 21 L.A.C.4th 223 (Man., Cherniack) at 235, 236-38.

¹⁹*Re Winnipeg (City) and C.U.P.E., Local 500* (1990), 12 L.A.C.4th 231 at 240. See also *Re Elgin Country Roman Catholic Separate School Board*, *supra* note 16, at 216-18.

²⁰*Supra* note 2, at 6-64.

The first requirement is relevance. The test also must relate to the requirements of the job:

A test that has no direct application to the job, but rather is intended to assist the company in ascertaining whether the applicant has an aptitude in a particular field, cannot be relied upon as a proper assessment of the knowledge, skill and efficiency of that applicant on the job. The test must be a reliable guide for the purpose for which it is used and not so generalized that it bears no reasonable relation to the available position.²¹

This basic standard for testing was set out by Arbitrator Weatherill in the *Polymer* decision as follows:

In many cases it is to be expected that an employer would construct and administer his own test, designed to reveal the ability of employees to perform the particular work he requires to be done. . . .

The proper purpose of this procedure was to assess the ability of the candidates for the posted job—that is, to enable a reliable and accurate prediction to be made as to their future performance. In our own assessment of the evidence, and in the opinion of the properly qualified experts called to testify in this case, the interview procedure which has been described simply did not justify such predictions.²²

The union's expert witness stated in her report:

The process of choosing an instrument for personnel decision-making starts with a thorough job analysis. This is an essential step that cannot be left out under any circumstances.²³

The most troublesome aspect of the process used to select for the position of Chargehand is the lack of a job analysis. The requirements that were outlined for the Chargehand position are very broad, and except for the journey person's status, are characteristics that would be required for most supervisory/management positions (leadership skills, self-motivation, extensive product knowledge, willingness to make a commitment to a team environment). Unless the selectors knew specifically the Knowledge, Skills, Abilities, and Other characteristics (KSAOs) necessary to adequately perform the Chargehand job, it is unclear how an effective selection decision was made.²⁴

The union's expert witness amplified in oral evidence the kind of detail that might be expected if a job analysis had been done on the position of chargehand. Neither the position specifications nor

²¹*Testing at Arbitration: An Arbitrator's Perspective* in Labour Arbitration Yearbook 1993 (Butterworths-Lancaster 1993), 177 ff.

²²Orsborn & Marshall, *Testing at Arbitration: A Management Perspective* in Labour Arbitration Yearbook 1993 (Butterworths-Lancaster 1993) 183, 188.

²³Kline, Review of Testing Procedure and Evidence at Grievance Arbitration, Feb. 2, 1995, at 3.

²⁴*Id.* at 14.

the job descriptions in the job postings could be considered a job analysis of the chargehand position.

The evidence showed that the employer did not proceed through the steps of first, an analysis of the actual requirements of the job, and then, construction or selection of tests to match those actual requirements. Rather, the employer chose tests designed for very different purposes, and purported to use them as a measure of how the candidates would perform in the particular position of chargehand. The employer's expert stated in examination in chief that he would not have designed that selection process for the employer. The employer could easily have carried out a job analysis, since it has the same or very similar positions in British Columbia, as acknowledged in evidence by the company's human resources (HR) manager.

The second dimension for review of tests is that of validity and reliability. The union's expert emphasized that the "reliability of instruments used in personnel decision-making should be 0.90 or more. . . ." In oral evidence, the expert explained that stringent standards must be expected when decisions affect people's lives. The employer's expert agreed with Nunnally and Bernstein's stringent standard of reliability when decisions can affect people's lives. In cross-examination, the representative for one of the testing companies agreed that any decision on a person's life should be handled judiciously.

The Wonderlic Personnel Test. There was no issue among the experts as to the reliability and relatively high validity coefficient of the Wonderlic Personnel Test. The union's expert noted the generally positive reviews of the test in the literature. The company's expert testified that it is a good test and has had "tons of validation" but that it was not one that this expert would use.

However, the union's expert highlighted a significant caution contained in the test's User's Manual: "While test scores are very powerful predictors of future performance, they are not sufficient to automatically outweigh other qualifications such as prior employment, educational achievement and personal references." The expert cautioned against overreliance on the test, or on a notion of "passing the test." Even the company's expert, a booster of cognitive testing, would not advise employers to rely on only a test result in personnel selection.

It was inappropriate for the employer to have eliminated any candidate from the process based on this test result alone.

The Common Sense Manager (TKIM). No expert at the arbitration hearing approved of TKIM. The union's expert noted that the reliability of the test was too low to warrant its use in a selection process and that the test is still in the developmental stage and should not be used yet in a selection process. The company does not use or recommend that particular test. Since the employer had not established the relevance, reliability, or validity of this test, the results from this source should not have constituted a justifiable component of the chargehand selection process.

The TKIM involves a comparison of the candidate scores to the scores of a management profile. The HR manager stated in evidence that the managers selected by the employer to constitute the manager profile were *senior* managers (branch managers, customer service managers, warranty managers), not first-line managers. The employer had available as a pool from which to draw a representative sample of first-line managers both the group of chargehands in British Columbia and foremen in Alberta. The employer could have constructed a comparison more directly related to the chargehand position but failed to do so.

Thomas International Personal Profile Analysis (PPA). The union's expert stated unequivocally that the PPA should not be used in the personnel selection process. The expert criticized the test as follows:

- It is not clear how the personality traits of Dominance, Influence, Steadfastness, and Compliance are related to successful performance in the chargehand position. That link has not been made.
- It is inappropriate to use it to disqualify an individual from a selection process. If it is used at all, it should be used only to clarify the expectations of the job in an interview.
- Individual profiles should not be compared to each other; one cannot compare across people.
- There are no external independent studies of the test. (The corporate representative acknowledged that there has not been as yet an independent study of the PPA.)
- The test's Handbook made clear that a job analysis is a necessary part of the process, and such was not done by the employer.²⁵

The employer's own psychologist stated in evidence that he would never use the PPA.²⁶

²⁵"Re: Review of Testing Procedure & Evidence at Grievance Arbitration" by Dr. Theresa Kline.

²⁶*Id.* at 12-13.

The test representative from the developer agreed with the company advocate's use of the PPA in the context of a selection procedure of seven stages. They do not refer to it as a "test," but rather recommend that the report generated by the test be used in a discussion with the individual. Its main purpose is as a useful mechanism in a structured interview, and it should be "used in a dialogue with the individual. It should never be used to preclude any individual from a position. We never say, 'Don't hire based on this instrument.' That is a central concept to the system." The representative agreed that the computer printout generated by the test should *not* supplant the full human process and that the final decision must be made by people in the context of many sources. Although the representative stated that the PPA should never be used to preclude any individual from a position, the HR manager for the employer stated in evidence that in fact the PPA was used to disqualify one of the grievors.

The test representative also stated that they recommend that the initial step of a job analysis (HJA) be done in all cases. The HJA, the representative explained, comes out of a subjective impression of people analyzing a job. Someone close to the job, perhaps three or four peers of but not the incumbent, should complete the HJA. If it is a new position, those reviewing the job should be in positions as close as possible to the position being analyzed. The employer, in this case, did not carry out the HJA, although it certainly could have done so. The HR manager stated in evidence that the chargehand positions in British Columbia were similar to what was contemplated for Alberta. The British Columbia chargehands, or persons in positions close to them, could have completed the HJA.

The PPA cannot be considered an appropriate component of the selection process for the chargehand position. The employer did not prove it to be valid or reliable. Further, the employer did not use it in the manner recommended by its developers. If any candidate was disqualified based on the test, such disqualification cannot stand. The employer was unfair and irresponsible when it generated or sanctioned reports without sound validation or reliability, which included sweeping statements about employees that could have profound and serious consequences for their work lives.

Minimum test scores

In *Re Domtar, Inc.*, Arbitrator Solomatenko ruled that "if a minimum score on a test is required, then such minimum must be

made clear by the Employer and must be applied consistently.”²⁷ The employer in this instance did not set out in advance the minimum expected in any of the procedure’s uses and did not convey an overall threshold. Further, the employer has been inconsistent among candidates. For example, the test scores of one grievor were not clearly different from scores of other candidates maintained on the shortlist, yet he was removed from the shortlist. Other candidates received test scores significantly higher than the candidates on the shortlist but were not put on the shortlist.

Remedy Requested

The union requested that, given the following considerations, the proper arbitral remedy for the grievances was to place the grievors in the position of chargehand according to seniority. The evidence established that:

- The grievors are clearly qualified to hold the position.²⁸
- The collective agreement directs that seniority is the prime factor in filling job vacancies, including promotion.
- The grievors performed successfully in comparable leadhand positions.
- The grievors have been subjected to unfair and unreasonable procedures by the employer.

The Award

The policy grievance filed by the union had in essence requested the following remedies:

1. Declare invalid all use of tests.
2. Declare invalid the results of the particular tests used.
3. Direct that the test results be struck from the files of all union members who took the tests.
4. Declare that Article 5 applies.
5. State what are the proper procedures for a selection process.

²⁷*Re Domtar, Inc. (Domtar Construction Materials) and United Electrical, Radio and Machine Workers, Local 561* (1986), 22 L.A.C.3d 247 (Ont., Solomatenko) at 249, 251–51.

²⁸*Finning Ltd. v. I.A.M., Local 99 and Dhillon and Lokos and Waye*, Alta. G.A.A. 95-012G (July 28, 1995) (Beattie).

In terms of remedy, we had convinced our client prior to the arbitration that in respect to the first remedy, based on the arbitral jurisprudence, we could not have succeeded in putting forward a position that testing is *never* acceptable, or that *no* test is valid. In terms of remedies 2 and 3, in relation to the particular tests, Arbitrator Beattie stated:

- [The Wonderlic] was the best assessment tool, but it should not be used in isolation;
- [The TKIM] should not be used by itself, if at all; the expert profile used was suspect;
- [The PPA] should not be used by itself, if at all; if used, it should be in conjunction with a Human Job Analysis and then only for providing information in an interview.²⁹

Arbitrator Beattie, however, did not go on to give any direction as to what should flow from his conclusions, in particular, that the test results be removed from the files of all union members who took them. The union was nonetheless able to achieve this through the letter of understanding.

In relation to remedy 4, Arbitrator Beattie was clear in the introductory summary and in the main text that Article 5 applies.³⁰ He failed, however, to address precisely how the interview circuit and tests fit within a threshold clause where the senior qualified person was entitled to the job.

In relation to the request for a statement as to proper procedures, Arbitrator Beattie gave a number of indications, some of them contradictory. He declared, for example:

- There was no job description.³¹
- The required qualifications were not stated clearly in the job posting.
- No appropriate minimum standard was established or consistently applied.
- Whether the TKIM and the PPA serve any purpose in a chargehand selection process is a significant concern.³²
- The interview ratings should not be deemed to have reliability.³³

²⁹*Id.* at 68.

³⁰*Id.* at 1, 58–61.

³¹*Id.* at 65.

³²*Id.*

³³*Id.* at 70.

But he also stated:

- The selection process was in principle reasonable.³⁴
- The interview process designed by the HR manager did contain a weighting of factors.³⁵

Arbitrator Beattie failed to address the union argument that there was an “over-weighting” of managerial factors, presumably accepting that this was a position different in kind as well as degree. He also rejected the union submission and supporting law that promotability could not be considered a proper requirement for the position.

The union and the employer may agree that they are left with many unanswered questions. It may be that the employer prefers it that way. If there was any doubt, there can be no further argument that the position must be awarded to the most senior qualified person. But beyond that blunt statement, how is the union to ensure objective and fair compliance with that hard-won seniority protection? Tests and interviews, we fear, will yet be used by the employer as instruments whose primary purpose is to dissipate to the point of irrelevance the mandated qualifications of Article 5. Training, knowledge, experience, and past performance, therefore, remain suspect. Training, experience, and past performance are particularly suspect as being somehow equated with seniority. Instead, reliance is placed on marks in cognitive tests administered to employees 22 years away from the interview circuit and 22 years into service with a company that now wishes to promote on the basis of a “g” factor.

Simple paranoia, predictable union resistance to change? The employer expert testified that 20 years on the job was simply doing the same job for 2 years 10 times over. That expert felt the minimum on the Wonderlic set by the test developers at 22 should more properly have been much higher, so high, in fact, that only 1 of the more than 50 heavy duty mechanics who wrote the test would have qualified. No problem, that expert witness responded, if there are no qualified candidates, go outside. In fact, if he had to choose, a single cognitive test would be his instrument of choice. Attempts to require that tests be related to the job, in that expert’s view, are just outdated legal restrictions. This is not controversial,

³⁴*Id.* at 69.

³⁵*Id.*

we were assured. The expert then cited a recent book entitled *The Bell Curve*.³⁶ These propositions, he suggested, will be generally accepted in five years.

The award, in the union's view, did not go far enough in clearly placing the use of the tests and the interview within a balanced process. Appropriateness, reliability, and validity of the test are aspects of only one element in a balanced process; the use of a structured interview is another. Where the parties have addressed their minds to the promotions procedure, the arbitrator's task is, as always, to go back to the collective agreement. The mandated qualifications are the true measure of a balanced and complete process. Where the parties have agreed upon a threshold qualification, a "g" mark on a cognitive test must be placed in the context of the mandated qualification, the requirement of a balanced approach, and the fundamental entitlement of seniority. It seems clear that the "g" factor is dissipated, not knowledge, training, experience, and past performance. A passage from *The Bell Curve Debate* is apt:

As Charles Darwin once wrote to his cousin Francis Galton, founder of the eugenics movement: "I have always maintained that, excepting fools, men [do] not differ much in intellect, only in zeal and hard work."³⁷

III. EXPERT PERSPECTIVE

THERESA J.B. KLINE*

General Selection Process

This section is based primarily on the sources acknowledged as the standards for practice in the area of personnel testing, which include the Society for Industrial and Organizational Psychology's Principles for the Validation and Use of Personnel Selection Procedures¹ and the Canadian Psychological Association's Guide-

³⁶Hernstein & Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (Free Press 1994).

³⁷Easterbrook, *Blacktop Basketball and the Bell Curve* in *The Bell Curve Debate* (Times Books 1995), 31.

*Associate Professor, Department of Psychology, The University of Calgary, Calgary, Alberta.

¹Society for Industrial and Organizational Psychology, Inc., *Principles for the Validation and Use of Personnel Selection Procedures*, 3d ed. (Society for Industrial and Organizational Psychology, Inc. 1987).

lines for Educational and Psychological Testing.² Selection procedures are based on the assumption that some important aspect of job behaviour can be predicted from a selection tool such as a written test or interview. It is widely acknowledged that any single test or battery of tests should comprise only one aspect of the entire selection system. There are many other sources from which to obtain valuable information about an applicant or employee, such as past work performance records, peer evaluations, education and training records, and past experience. All of these, and other selection/assessment tools, should rightfully be part of the selection system when appropriate.

The process of choosing an instrument for personnel decision making starts with a thorough job analysis. This is an essential step that must not be left out under any circumstances. The job analysis provides the organization with information concerning important activities performed on the job. From this list of activities, the knowledge, skills, abilities, and other characteristics (KSAOs) deemed to be necessary for successful job performance can be derived. This information is essential for the development of relevant performance criteria and for making informed judgments about specific instruments chosen and developed for the selection process. Without a clear idea of the job one is selecting for, it is fruitless to search for an instrument that will predict successful job performance.

The next step in the personnel decision making process is to determine which predictors, if any, are related to job performance. These techniques are subsumed under the general heading of "validation." It should be noted that tests themselves are not validated. Instead, the specific uses of the test are validated.

There are three broad strategies that one can use for validation purposes. The first is the criterion-related strategy. In this strategy, the relevant criteria (job performance as measured by supervisor ratings, for example) and the relevant predictors (personnel tests, for example) are identified. The higher the co-relationship between the criterion (job performance) and predictor (test), the more valid the test.

The second strategy is the content-oriented strategy. In this strategy, the job analysis information is critical. The job analysis process identifies the important job tasks, behaviors, and KSAOs

²Canadian Psychological Association, *Guidelines for Educational and Psychological Testing* (Canadian Psychological Association 1987).

needed to perform successfully on the job. Next, tests, test items, or interview questions are analyzed by "subject matter experts," which usually include position managers, personnel professionals, and job incumbents. These individuals are in a good position to determine the degree of congruence between the job tasks, behaviours, and KSAOs identified through the job analysis, on the one hand, and the content of the particular tests, test items, or interview questions, on the other.

The third strategy is the construct-oriented strategy. In this strategy, the specific abstract concepts (i.e., constructs) such as personality characteristics and intelligence that are deemed to be important determinants for job success are determined. These were identified through the job analysis procedure. The next task is to find an instrument that purports to assess the constructs of interest (e.g., intelligence). An assessment is then made of the instrument to examine the extent to which it actually is an effective measure of the construct. If it is an effective measure, then this instrument can be used in the selection process.

It should be noted that these validation strategies are not mutually exclusive. That is, a combination of strategies may, and perhaps should, be employed for validation purposes whenever feasible.

Another critical feature of any selection instrument is its demonstrated reliability. The reliability of an instrument is an indication of its consistency. There are several types of consistency that should be demonstrated by any selection instrument. One is consistency over time (test-retest reliability). In this type of reliability, a score an applicant receives on a test at one time, and the score that he or she obtains at a later time on the same test is assessed for consistency. Another common type of reliability is interrater reliability. In this type of reliability, the ratings of an applicant made by several judges are assessed for their similarity.

A Test of General Cognitive Ability

This type of test measures general cognitive ability in the work domain. The particular test involved in this arbitration is a 50-item, 12-minute test first published in 1938 that is representative of a number of cognitive tests. The test measures verbal aptitude, quantitative aptitude, and spatial aptitude. A note in the user's manual of the particular test used states that the test does not measure how well a person will employ his or her ability.

The basic rationale for using a general cognitive ability test in the personnel selection process is that all jobs require some degree of mental skill (i.e., cognitive ability). Extensive reviews of the literature have indicated that cognitive ability is a valid predictor of job performance.³ Performance in jobs requiring more mental skill is better predicted by this test than in jobs that do not require a high degree of cognitive functioning.

To demonstrate that this test has construct validity, it is necessary to show that cognitive ability is adequately captured by the test. Scores on this test are correlated to scores on other, longer, established intelligence tests, so that it is reasonable to conclude that it does assess general cognitive ability. Reviews of this test by experts in the field are very positive. The test-retest reliabilities are high, and the extensive norms provided in the manual are useful. The criterion-related validity coefficients between scores on this test and subsequent job performance are relatively high for both middle managers and first-line supervisors.

One of the criticisms voiced has been that the minimum cutoff scores for various occupations are not justified as fully as they should be. As noted in the manual itself, there is no such thing as a true minimum score; there will always be some successful applicants that score below any chosen minimum. This is the reason for avoiding the use of any single index as the sole basis for a personnel decision. In fact, the user's manual makes this very clear by stating that test scores are very powerful predictors of future performance, but they are not sufficient to automatically outweigh other qualifications such as prior employment, educational achievement, and personal references. In addition, one reviewer of the test specifically indicates that minimum or typical scores on the test for various occupations are likely to be misleading to many users. There is a risk that users will adopt these standards without sufficient substantiating information or with only a superficial basis for matching jobs with listings in the tables.

The conclusions one should draw from the existing literature on this test are that: (1) it is a reasonable measure of mental ability; (2) the more cognitively complex the job, the better that tests such as the one used in this case will predict job performance; (3) not all of the variance in job performance is accounted for by tests of mental ability; thus, scores on the test should be used as part of the

³Schoenfeldt, *Review of the Wonderlic Personnel Test* in *The Ninth Mental Measurements Yearbook* (Buros Institute 1985), 1757-58.

decision making process, but one should not make a personnel decision based solely on test scores; and (4) caution should be used in adopting the general minimum cutoffs listed in the user's manual for any specific job.

A Specific Knowledge Test

This type of test was designed to identify individuals whose "street smarts" indicate that there is a high potential in them for excellent performance in managerial and executive careers.⁴ The particular test used in this case presents test-takers with nine scenarios and questions asking what they would do in the situation presented. They rank-order the choices presented, and that ranking is compared with a standard. The closer the applicant's choice is to the standard, the better the score.

The rationale for using an instrument such as this is that knowledge is not always taught formally; it is gained through informal mechanisms and is characterized as practical know-how. The amount of this tacit knowledge that an individual possesses is a key determinant in effective managerial performance. The construct of tacit knowledge has recently come under fire from the academic community. The critics of tacit knowledge argue that the construct of tacit knowledge does not exist, but rather it is subsumed under general cognitive ability and specific job knowledge.⁵ In other words, tests of tacit knowledge have nothing more to offer than do tests of general mental ability and tests of specific job knowledge. In addition to the theoretical criticisms, some reviewers of the test have argued that there are several methodological concerns about the data presented in support of such a test.

I would agree with the critics. The construct of tacit knowledge has yet to be accepted by the academic community. In the few published studies available, the findings suggest that the positive results about the use of the test that had been presented in previous studies were not meaningful for female managers,⁶ and the method

⁴Sternberg & Wagner, *The G-centric View of Intelligence and Job Performance Is Wrong*, 2 *Current Directions Psychol. Sci.* 8-9 (1993); Wagner, *Tacit Knowledge in Everyday Intelligent Behavior*, 52 *J. Personality & Soc. Psychol.* 1236-47 (1987); Wagner & Sternberg, *Tacit Knowledge in Managerial Success*, 1 *J. Bus. & Psychol.* 301-12 (1987).

⁵Jensen, *Test Validity: g versus "Tactic Knowledge"*, 2 *Current Directions Psychol. Sci.* 9-10 (1993); Ree & Earles, *g Is to Psychology What Carbon Is to Chemistry: A Reply to Sternberg and Wagner, McClelland, and Calfee*, 2 *Current Directions Psychol. Sci.* 11-12 (1993); Schmidt & Hunter, *Tacit Knowledge, Practical Intelligence, General Mental Ability, and Job Knowledge*, 2 *Current Directions Psychol. Sci.* 8-9 (1993).

⁶Horgan & Simeon, *Gender, Mentoring, and Tacit Knowledge*, 5 *J. Soc. Behavior & Personality* 453-71 (1990).

of calculating the scores (comparing actual to ideal scores) was not a useful way to evaluate performance on the test.⁷

The reliability of this test is too low to warrant its use in decision making.⁸ The criterion-related validity coefficients reported were not bad, but the criterion variables used were questionable measures of manager performance. They included such things as salary level, whether or not the manager was in a Fortune 500 company, new business generated, and policy implementation. Certainly, specific managerial skills such as communication, leadership, and administrative skills have yet to be shown to be predicted by scores on the test.

The user manual indicates that it is necessary to first establish a manager profile from within the organization, and this should be based on responses of at least 30 managers in the organization. Once the profile is created, an applicant's score is assessed in relation to the profile. This profile forms the basis on which one should make a decision about the potential management match of the particular applicant to the specific organization. It should be noted that a reliability test should be carried out on the expert responses to determine the extent to which the existing managers' responses agree. There is no evidence that the organization involved in this particular case established a profile with which to compare the applicants' scores.

This test is clearly still in the development stage. There needs to be many more studies to show that: (1) tacit knowledge is indeed a construct separate and distinct from general cognitive skill and job knowledge; (2) the reliability of the test has been brought to a substantially higher level than at present; and (3) test scores are related to well-defined management performance indicators. This research should be conducted before determining that the test is worthy of inclusion in a personnel decision making process. On the basis of the previously presented information, it is my professional opinion that the test is not yet ready for use in a selection context.

A Personality Test

The particular personality test involved in this arbitration is more of an aid to understanding the motivation of human behav-

⁷Kerr, *Tacit Knowledge as a Predictor of Managerial Success: A Field Study*, 27 Canadian J. Behavioural Sci. 36-51 (1995).

⁸Nunnally & Bernstein, *Psychometric Theory* (McGraw-Hill 1994).

ior than it is a selection device. It is best characterized as providing information about an individual's self-image. This test claims to be a measure of four personality traits: dominance, influence, steadfastness, and compliance. These traits are measured by having respondents indicate which of four presented characteristics is "most like" them, and "least like" them. Respondents do this for 24 sets of four characteristics. The theory postulates that the "most like" selection is made by the testee who is viewing the self as the individual is viewed by others. The "least like" selection reflects the self under pressure, and the summary of the "least like" and "most like" selections is interpreted as the testee's self-perception. The candidate's profile is compared with a profile of what would be expected in the job. The more congruent those profiles, the better.

Presumably, the reason for using a personality test in the personnel selection process is that there is a relation between personality and performance on the job. This should be determined from the job analysis procedure. It is not clear how the personality traits measured by this test are related to successful performance in the position in question. There are some recently reported examples of how difficult and time-consuming a process it is to determine the specific personality traits necessary for successful performance on a specific job.⁹

Two technical reports¹⁰ on the test, which were commissioned by the company that publishes and markets the test, indicate that the reliability of the test is fairly high and that the four personality dimensions assessed by the instrument seem to be distinct one from another. The validation evidence of the usefulness of scores on this test as a selection device is mixed. There are some reasonably high criterion-related validity coefficients and others that are very low. The technical report admits that it is impossible to predict success in a complex operation on the basis of individual scores on the test. It also states that use of the test as a basis for selection will not guarantee improved performance. A technical handbook that

⁹Day & Silverman, *Personality and Job Performance: Evidence of Incremental Validity*, 42 *Personnel Psychol.* 25-36 (1989); Gough, *A Managerial Potential Scale for the California Psychological Inventory*, 69 *J. Applied Psychol.* 233-40 (1984); Hogan & Hogan, *How to Measure Employee Reliability*, 74 *J. Applied Psychol.* 273-79 (1989); Hogan, Hogan & Busch, *How to Measure Service Orientation*, 69 *J. Applied Psychol.* 167-73 (1984).

¹⁰Hendrickson, *Personal Profile Analysis: A Technical Manual* (Thomas International Systems Europe Ltd. 1983); Irvine, Mettam & Syrad, *Valid and More Valid? Keys to Understanding Personal Appraisal Practices at Work*, 13 *Current Psychol.* 1, 27-59.

accompanies the test¹¹ indicates that the test should be used as an aid in the interviewing process. It should be used to help in structuring questions to find out more about the candidate in an interview situation.

I found not one reference to the test in the PSYCHLIT citation index, which accesses approximately 1,500 journals in all areas of psychology, as well as communication, education, mental health care services, biology, and applied psychology. The claim made in one report of the test¹² that it is a splendid tool for team building, selection, placement, career guidance, and for assisting in resolving interpersonal conflict is absurd. At best, this test is in the infancy stages of being developed as a personnel selection tool. Personality profiles should not be used for screening candidates in the personnel selection process. They should instead be used to help guide the interviewing process.

An Evaluation of the General Selection Process Used

The most troublesome aspect of the process used to select for the position in question was the lack of a job analysis. The requirements that were outlined for the position are very broad, and, except for the journey person's status, they are characteristics that would be required for most supervisory/management positions (leadership skills, self-motivation, extensive product knowledge, and willingness to make a commitment to a team environment). Unless the selectors knew specifically the KSAOs necessary to adequately perform the job, it is unclear how an effective selection decision could be made.

A second concern is with the use of the test scores in the selection process. I had been told that an initial shortlist was drawn up based on the interviews and that this shortlist was further pared down on the basis of test scores. However, there was inconsistency in the application of the cutoff scores used; some candidates were retained that had lower scores than those that did not make the shortlist.

A third concern is with the use of the personality test. As noted in the previous section, the use of the information gained from the

¹¹Irvine, *Personal Profile Analysis: Technical Handbook: A Contemporary Frame of Reference for Research and Development* (Thomas Lyster Ltd. 1988).

¹²*The Kaplan Report* (Kaplan Associates 1991).

test is highly suspect in the present context. That is, scores were used to make personnel decisions, an action that the test publishers warn against. In addition, it is noted in all of the test results that: "As we have not been given a Human Job Analysis for this post, it has not been possible to make a detailed assessment of (applicant name)'s strengths and weaknesses against the specific job requirement." Again, the need for a clear statement of the KSAOs necessary for the job (i.e., a job analysis) is highlighted. The test publishers indicate that applicants' profiles should serve as a point of discussion regarding how well they will "fit" the specific job. In addition to the difficulty with using this particular test, most personality tests tend to be very poor predictors of job performance. The reason for this is that "personality" is a very broad thing and is best related to general behaviours. Job performance, at any given time, is only a very specific aspect of peoples' lives, and general personality traits tend to be unable to predict specific behaviours at specific points in time.

A fourth concern is with the lack of interrater agreement between the five raters on many of the interview ratings. The seven dimensions of Technical Expertise, Communication Skills, Management Skills, Administration Skills, Leadership Skills, Quality, and Personal Attributes were assessed in the interview and rated on a scale from 1 to 5 (1 = poor, 2 = below average, 3 = average, 4 = above average, and 5 = excellent). For one grievor, the raters varied in their responses substantially, with Communication Skills ratings from 2–4 (below average to above average), Management Skills ratings from 1–3 (poor to average), Administration Skills ratings from 1–3 (poor to average), Quality ratings from 2–4 (below average to above average), and Personal Attributes ratings from 2–4 (below average to above average). The Technical Expertise ratings from 4–5 (above average to excellent) and the Leadership Skills ratings from 2–3 (below average to average) were much more consistent. For another grievor, the raters also varied in their responses, with Management Skills ratings from 0–4 (less than poor—zero is not even on the scale—to above average), Administration Skills ratings from 1–4 (poor to above average), Leadership Skills from 2–4 (below average to above average), and Quality ratings from 3–5 (average to excellent). Technical Expertise ratings from 3–4, Communication Skills ratings from 3–4, and Personal Attributes ratings from 3–4 (all average to above average) were much more consistent. These interviewers exhibited very low to nonexistent reliability.

Regarding the overall selection process used by the company in this case, my conclusion is that there should be an attempt to make the process more rigorous. First, a thorough job analysis should be conducted. Second, written tests, interviews, or other selection tools such as work samples that specifically assess the KSAOs needed to perform the job should be chosen with great care or developed in-house. Third, raters should be trained in the areas of (1) consistency in what to look for on each of the relevant dimensions assessed in the interview process, and (2) consistency in what each of the values (1 to 5) on the rating scale represents.

IV. EXPERT PERSPECTIVE

SEAN MACENNIS*

This paper is intended to be practical, while recognizing the relevance of a theoretical understanding. I will begin by highlighting three guiding principles in test content, then turn to the intensely practical question of the influence of equal employment opportunity regulation on hiring practices. With both research and practical issues as the major focus, I will then present several hypotheses based on real time experiences.

The content and selection of any test used in the recruitment process must be dictated by three broad principles. The first is that *the test be defined clearly in the literature by published works that reveal those aspects of human mental performance, behavior, and personality captured by the test items*. Theory must prescribe in practice what specific qualities the items demand of people and what makes these demands progressively more difficult to fulfill.

The second principle is that *it have a long history as a type of test that measures known abilities used in work, training, and educational contexts*. The tests used were, therefore, planned to demand from applicants the following essential mental and behavioral qualities: constant attention and concentration, memory for task procedures, accuracy of decisionmaking, speed of processing information, and relevance of identified job-specific behaviors.

The third principle is the test's *equal opportunity knowledge requirement or threshold*. School graduates may be expected to process information and in so doing to know the alphabet and the order

*President, Thomas International Management Systems (Canada) Ltd.

of its letters; to recognize letters in upper and lower case versions; to compare simple adjectives like heavier and lighter, gentle and humble, to use negatives to change the meanings of simple sentences; and to count up to 30 and to subtract two numbers not greater than 30. What is now commonly referred to in testing literature as "Big Five Theory" demands that any recruitment test be built around these standard content thresholds, so that the mental and behavioral qualities tested are not inhibited by special or advanced educational requirements.

The acquired knowledge demands of any test or assessment vehicle are, therefore, no more than functional literacy and numeracy. Such a test also may be described as minority-conscious, that is, it is constructed in such a way as to minimize the effects of educational disadvantages.

Selection decisions based on tests and inventories could be termed pointless if it were not for the fact that someone must make a judgment about who should be hired. Judgment is an integral part of most, if not all, selection procedures. If performance evaluations, peer evaluations, and education and training records are used for promotions (or as criteria in empirical research on other predictors); if past experience and work samples are rated and scaled by observers or inspectors; if history of promotions or salary growth are used; if interviews, assessment centers, or personal appraisals are used; if someone or some group of people are given the responsibility for considering and weighing available evidence and presumably basing a selection decision on that evidence; if any of these is part of the recruitment and selection process, then judgment is present. The question is not whether there is subjectivity in selection decisions but whether that subjectivity is recognized, reliable, and *understood*.

The idea behind equal opportunity is a simple one: there are human predispositions that tend to put whole classes of people at an unfair disadvantage in the job market. To help offset such biases and enable all people to compete fairly and equally, an organization's recruitment and selection process must meet the overall intention of the Equal Employment Opportunity Commission (EEOC) code as it applies to the prohibited grounds of discrimination. Title VII of the Civil Rights Act of 1964 explicitly permitted employment tests, except where used to discriminate on the basis of race, sex, religion, or national origin. Documents issued later were influenced little by experts in psychometrics or selection research. The *Uniform Guidelines on Employee Selection*

*Procedures*¹ is the most recent text used as a foundation for debate on this topic under discussion. Key issues that must be addressed include: What questions may and may not be asked during the interview? What is the purpose of each question? What is the role of testing? How should this information be applied?

Clearly, any organization has the right to set its own selection criteria to determine suitability of a candidate for the position. These criteria, however, should not be discriminatory, nor should the questions asked be construed as discriminatory. The following key points are notable with regard to testing: if the test has not been properly validated and checked for reliability, it will be regarded as discriminatory; if the test is undertaken by persons who are not qualified to administer or interpret such tests, it will be regarded as discriminatory; research is required to show that adequate norm tables are available and used to take into account cultural and other differences; it is important to be able to practically demonstrate that the characteristic that the test measures is an important characteristic required for job success.

In seeking evidence of equal opportunity compliance, psychometric validity, or job relatedness, the whole process traditionally has been called *validation*. This is a "loaded" term and is abused freely by experts and test suppliers alike. Although the term "validation" implies approval, a more appropriate term would be more neutral such as "investigation" or "evaluation." The implication rests on building an argument for or against a particular use.

Robert M. Guion argues that the term "validation" has been burdened with so much excess baggage that he prefers the more neutral term "evaluation."² He contends that the basic question of validity is how well the intended attribute has been measured. No single definitive answer is possible, in his opinion. Guion offers seven broad categories "in the spirit of guidance" in evaluating psychometric validity:

1. Is the measurement procedure based on a formal theory of measurement?
2. Was there a clear idea of the attribute to be measured?
3. Are the mechanics of measurement consistent with the concept?
4. Is stimulus content appropriate?
5. Was the instrument carefully and skillfully developed?

¹U.S. EEOC, Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. 1607.

²Guion, *Personnel Assessment, Selection, and Placement*, in *Handbook of Industrial and Organizational Psychology*, 2d ed., vol. 2, eds. Dunnette & Hough (Consulting Psychologists Press 1991), 375-78.

6. Is the internal statistical evidence satisfactory?
7. How do the scores correlate with other variables?³

Guion explains that the above questions “and the kinds of evidence they imply, are neither complete nor detailed, but they can confuse the bureaucratic ‘stamp collector’ who prefers to match a ‘kind’ of validity and a specific ‘strategy’ for validation.”⁴ I would suggest that with well-meaning and substantive attempts at test “validation,” where a body of recognizable published evidence is openly available and where the architecture of such tests is open to continual improvement, these tests are given their due within a pooled judgment environment.

From the fast, radically shifting business environments of today, I feel confident in suggesting six additional items for the practically minded person tasked with the responsibility of selection:

1. *Convenience.* Whatever the context, small or large groups, the tests are easy to administer, score, and interpret.
2. *Economy.* Low costs of upkeep and renewal are ensured because new forms of the tests need little or no adjustments to maintain equivalence.
3. *Parallel forms.* The almost unlimited number of parallel forms of the items that can be generated provide the user with some important operational choices that conventional tests are unable to offer.
4. *Low educational threshold.* The low threshold of educational skills required to perform the tasks give the tests themselves an extended range of uses.
5. *Delivery options.* By basing the construction of tests in microcomputers, we are able to provide not just many versions of the same test but to offer a choice in the delivery of the tests to users.
6. *Support.* Skilled help must be available for any pilot study or operation of tests and tasks. A mandatory period of training is followed up with regular workshops to evaluate practical applications from the field.

In general, I would recommend that exploration of options, costs, and benefits for users should be the first priority. Second, I would not advise the use of tests without planning to evaluate their use in any particular context through appropriate reliability and

³*Id.* at 375.

⁴*Id.*

validity studies. A common term used in relation to this is "benchmarking."

The nature of business is fluid, and within this simplified context dwell all the elements that allow a business to function. Recruitment and selection are pieces of the business puzzle. Someone in an organization must decide whether the evidence to support a hiring decision based on job relatedness is valid enough to start to use a proposed predictor or battery of tests. If this person is qualified, has a solid understanding of the job, understands the logic of any predictive hypotheses involved, is familiar and assured of the validity of any measurements used, this person is qualified to decide, even without empirical evidence. Today, there is an excellent argument for involving pooled resources in making the selection decision. Quality of judgment may be sacrificed for consensus, but the ability to debate and reason, based on observable and identifiable job-based criteria, may increase reliability. Another benefit is that no one person need bear the full responsibility for the decision.

A common trait uncovered in business selection is the tendency to self-select based on one's personal affinity for a particular candidate. The intent may be innocent, but the outcome may not address the correlation of all available evidence for job suitability. It is quite clear that the objective of individuals tasked with selection is to understand the literature on individual differences in abilities, personality, and background, and among other considerations, to make judgments wisely about appropriate predictor constructs, not merely how to select, buy, and administer a test. Guion calls for more ongoing research into personnel practices, and this bodes well for those of us whose professions revolve exclusively around recruitment and selection. The cost of a bad hire to organizations is well-documented; a multinational client in the staffing services business has published the cost at a modest number of 2.5 times annual salary.⁵ This same organization has produced empirical data to show an over 25 percent reduction in employee turnover over a three-year period as the result of a more systematic recruitment process.⁶ Similar examples are readily available. Organizations need to master the nuances of responsible recruitment processes without incurring penalties for nonadher-

⁵Kelly Services, Inc.

⁶Kelly Services (Canada), Inc.

ence to overzealous, scholastic interpretations of recruitment practice.

It is my belief that within a "best practices" framework, a reality check is needed that recognizes empirical research and science existing in harmony with the fast-paced practicalities of today's business environment. This requires a shared responsibility between business and academia that is nonpunitive and that recognizes the difficulties facing each. Contrasting expectations serve only to limit creativity and slow productivity. The goal should be *balance*, and in the words of Guion, "The days of personnel research have not ended."⁷

⁷*Supra* note 2, at 389.