

## **Validity Study**

# **Predicting the Job Performance of Maintenance Workers using a Job Knowledge Test and a Mechanical Aptitude Test**

Lester L. Sapitula & Marc C. Shartzler  
Los Angeles Unified School District

*Test scores from a Job Knowledge Written Test (JKWT) and the Wiesen Test of Mechanical Aptitude (WTMA) were correlated with job performance ratings of 102 maintenance workers. The results indicated no significant relationship between performance ratings and either the job knowledge test ( $r=.07$ ) or the ability test ( $r=.05$ ). An analysis of data from 782 applicants indicated that male applicants scored higher than did female applicants and white applicants scored higher than did minority applicants. However, the WTMA resulted in smaller race differences in test scores than the JKWT.*

---

### **Sample**

The validity study was based on 102 maintenance workers employed with a large metropolitan school district located in Southern California. Thirty-seven percent were Hispanic, 34% were White, 14% were Black, and 3% were Asian (12% did not report their race). Ninety-seven percent were male and 1 % was female (2% did not report their gender). Sixty-two percent were under 40-years old and 23% were over 40-years old (15% did not report their age).

### **Predictor Information**

Two predictors were involved in the validity study: (1) A Job Knowledge Written Test (JKWT) developed by school district staff involved in personnel selection and test development, and (2) the Wiesen Test of Mechanical Aptitude (WTMA) (Applied Personnel Research; Newton, Massachusetts).

The JKWT is a 95-item content-validated paper-pencil test designed to measure five job knowledge areas identified through a job analysis as essential for successful job performance: (1) Basic Knowledge of Hand Tools, Hardware, and Construction Crafts, (2) Loading and Moving Equipment, (3) Understanding and Following Instructions, (4) Arithmetic, and (5) Safety Practices and Procedures. The JKWT has a test-retest reliability of .61 with an 8-month lag between test administrations and an internal reliability (Cronbach's Alpha coefficient) of 0.88.

The WTMA is a 60-item paper-pencil test designed to measure a person's ability to understand and apply basic mechanical principles. Test items used images of common everyday objects (e.g., can opener, flashlight, and ladder) and reflect broad classes of physical and mechanical principles (e.g., basic machines, center of gravity, and transfer of heat). The WTMA has a test-retest reliability of .77 with an 8-month lapse between administrations and an internal reliability (Cronbach's Alpha) of 0.73. The correlation between the JKWT and the WTMA is .61.

### **Criterion Information**

The criterion measure was a 30-item job performance appraisal (PA) developed specifically for the validity study. The PA was designed to measure seven job performance dimensions that reflected the essential knowledge, skills, and abilities (KSAs) assessed in the JKWT (each job dimension consisted of three to six items). The seven job performance areas were (1) Use of Manual Tools, (2) Use of Power Tools, (3) Loading, Unloading, and Moving Equipment/Supplies, (4) Mathematical Ability, (5) Following Directions, (6) Safety, and (7) Performing Semi-skilled Work/Making Repairs. An incumbent's performance level on each item was evaluated using a 5-point rating scale ranging from 5 = *Exceptional* to 1 = *Unacceptable*. The internal consistency reliability analysis resulted in a Cronbach's Alpha coefficient of .96.

Before calculating the PA composite score, missing data found in the PAs were addressed. First, missing data within a dimension were estimated by calculating the mean of the incumbent's existing ratings within that dimension. Second, missing data for an entire dimension were estimated by calculating mean substitutions based on the sample population. Once the data manipulations were completed, the ratings in each dimension were summed. Then, the dimensions were factor-weighted based on the proportion of written test items measured in the JKWT. Finally, scores from each dimension were totaled resulting in a final PA composite score.

### **Validity Information**

There was no significant relationship between either written test and job performance. The validity coefficients for the JKWT and the WTMA were .07 and .05, respectively.

### **Limitations**

First, the researchers were instructed to design a PA that closely resembled the content of the JKWT. Therefore, a KSA-based PA was developed. Perhaps a behavioral-based PA would have yielded different results. Second, the researchers were working with archival data. This led to many research design and data analysis problems that could have been addressed before test administration. Finally, the PA data set required manipulations to estimate missing data. This practice decreases the variance in the data, which reduces the ability for a statistical test to discover relationships in a data set (Roth, 1994). As a result, the data manipulations may have also led to non-significant results.

### **Exploratory Analysis of Sex and Race Differences**

Group differences in written test scores based on race, gender, and age were explored using the entire applicant population who took both written tests. Note that within one year, both written tests were administered twice. Therefore, for ease of data analysis, the following results exclude data from repeat applicants.

*Sample.* The written tests were administered to 782 applicants. Forty-five percent were Hispanic, 28% were Black, 10% were White, 5% were Asian, and less than 1% was American Indian or Alaskan Native (11% did not report their race). Ninety-two percent were male and 4% were female (4% did not report their gender). Fifty-eight % were under 40 years old and 27% were over 40 years old (15% did not report their age).

*Data Analysis.* Using an alpha level of .05, regression analysis was employed to detect any group differences. Regression analysis seemed appropriate because of the unequal sample sizes within each group. Any significant results were not surprising because statistical tests can detect significant differences when the overall sample size is large. Therefore, it was necessary to calculate the effect size for any significant group difference in each written test in order to assess whether practical and meaningful differences existed. Table 1 summarizes the descriptive statistics, and Table 2 summarizes the effect sizes for each significant group difference in each written test.

**Table 1.**  
**Descriptive Statistics for Job Applicants**

<i>Demographic</i>	JKWT		WTMA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Race				
Asian (n=39)	55.6	10.1	42.9	6.1
Black (n=219)	50.8	11.3	40.2	6.4
Hispanic (n=352)	55.8	11.8	42.9	5.6
White (n=78)	68.4	11.6	46.4	5.7
Gender				
Female (n=31)	42.0	11.6	35.9	5.7
Male (n=719)	56.3	12.4	42.7	6.1
Age				
Under 40 yrs old (n=454)	55.5	12.0	42.6	6.2
Over 40 yrs old (n=211)	56.3	13.9	42.1	6.3

Note that descriptive statistics were not reported for American Indian/Alaskan Native because of the extremely small sample.

**Table 2**  
**Summary of Effect Sizes for Each Significant Group Difference in Each Written Test.**

<i>Group Comparisons</i>	Effect Size	
	JKWT	WTMA
White-Hispanic (n=430)	1.07	0.62
White-Black (n=297)	1.54	0.98
White-Asian (n=117)	1.15	0.58
Hispanic-Black (n=571)	0.43	0.44
Male-Female (n=750)	1.16	1.12

*Results on Race.* Significant race differences were found for the JKWT,  $F(4, 696) = 31.93, p = .000$ , and the WTMA,  $F(4, 690) = 14.78, p = .000$ . There were significant differences between White and Hispanic, White and Black, and White and Asian. Furthermore, there was a significant difference between Hispanic and Black. As shown in Table 2, with the exception of Hispanic-Black differences, the JKWT seemed to result in larger effect sizes than the WTMA, suggesting that the JKWT had larger race differences in test scores than the WTMA.

*Results on Gender.* Significant gender differences were found for the JKWT,  $F(1, 750) = 36.59, p = .000$ , and the WTMA,  $F(1, 744) = 28.180, p = .000$ . As shown in Table 2, it appears that both written tests yielded large effect sizes, suggesting that both written tests had large gender differences in test scores.

*Results on Age.* No significant group differences were found for the JKWT,  $F(1, 660) = .612, ns$ , and the WTMA,  $F(1, 654) = .781, ns$ . Thus, it appears that both written tests did not yield any significant age differences in test scores.

### **Limitations**

These were only results from exploratory analysis. Further research on actual group differences should be conducted. Additionally, some of the sample sizes in each group were grossly unequal, which may have skewed the results.

### **Recommendations**

Although there were no validity evidence for either written tests, the researchers recommend that the WTMA be used for future testing because it appeared to have produced smaller race differences (i.e., adverse impact would be minimized) in written test scores. Efforts to obtain validity evidence for both written tests should continue. Finally, the WTMA should be employed until validity evidence can be gathered for it and the current JKWT, or a new written test that produces significant validity evidence is developed.

### **References**

- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*, 537–560.
- Wiesen, J. P. (1997). *Technical Manual for the Wiesen Test of Mechanical Aptitude* (WTMA). Newton, Massachusetts: Applied Personnel Research.

### **Acknowledgements**

The authors would like to thank Dr. Joel Wiesen for allowing them to include the WTMA as part of the validation study at no cost to the school district. They would also like to thank the following school district staff: Dr. T.R. Lin, Assistant Personnel Director–Selection, who provided support and guidance throughout the process; Dr. Virginia Austin, Senior Personnel Analyst, who conducted the job analysis; and Doug Dunivan, Deputy Director–Maintenance, Maintenance & Operations Branch, who gave us permission to gather performance appraisal data from his staff. The authors currently work for the County of Los Angeles–Department of Human Resources and are involved in selection research and development.