

Information Exchange Article

A Criterion-Related Validation Study of Transit Operators

Greg A. Chung-Yan and
Peter A. Hausdorf
University of Guelph

Steven F. Cronshaw
University of Guelph & Organization and
Management Solutions Inc.

This paper presents the results of a theory-guided validation study conducted in a large urban transit authority. Four selection predictors were evaluated: an eye-hand-foot coordination test, a cognitive and perceptual ability test, a work preference inventory, and a behaviour description structured interview that assessed adaptive and interpersonal skills. These predictors were evaluated against a comprehensive domain of job performance that included on-the-job observations, accidents, complaints, commendations, absences, organizational commitment, and job and work involvement. Although the results were mixed, the results generally support the use of measures of non-intellective variables (i.e., work preference inventory and interview) when predicting transit operator performance.

Purpose

This paper presents the results of a concurrent validation study in a large Canadian transit authority for the job of Transit Operator (i.e., bus drivers). Results were used to recommend an operational selection system for the transit authority management.

Method

Sample and Procedure

After conducting a needs analysis with management to establish the objectives to be achieved in the test validation project, a thorough job analysis was conducted on the job of Transit Operator at the transit company. The primary job analysis method used was Functional Job Analysis or FJA (Fine, 1989; Fine & Cronshaw, 1999).

A sample of transit operator incumbents participated in the study on a voluntary basis during company time. The study was advertised to all transit operators with management cooperation. Before participating, all the volunteer

operators completed informed consent forms that included permission to anonymously observe and rate operator behavior at an unannounced time pursuant to collection of criterion data as part of the In-Service Riding Check (ISRC). There were logistical and practical constraints that prevented administration of all predictors and criteria to all operators ($N = 372$); therefore, not all validation analyses are conducted on the entire transit operator sample. For example, there were cost limitations on the number of operators who could be interviewed (125 received the Performance Oriented Structured Interview) and on the number of ISRCs that could be done (172 operators received scores on the Riding Check). These smaller sample sizes are reflected in the validation results for some of our predictors reported later.

Subgroup Membership

To ensure compliance with Canada's human rights and employment equity legislation, the study participants' subgroup membership was recorded. The subgroup classifications under this legislation are Majority male, Aboriginal, Visible minority, and Female. Majority males are defined as White Caucasian men. Aboriginals are defined as "people of Indian, Inuit, or Metis descent" (Human Resources Development Canada, undated, p. 21); Visible minorities are defined as "non-Caucasians in race or

*Address for correspondence: Greg A. Chung-Yan, Department of Psychology, College of Social & Applied Human Sciences, University of Guelph, Guelph, Ontario, Canada N1G 2W1. E-mail: gchungya@uoguelph.ca

non-white in colour”, but excluding Aboriginal persons (Human Resources Development Canada, undated, p. 24). Females do not include visible minority women.

Predictor Variables

Eye-Hand-Foot Coordination (EHFC) (Microcomputer Evaluation and Screening Assessment). The developers of the EHFC test define EHFC as the “[a]bility to move the hand and foot coordinately with visual stimuli” (VALPAR, undated). The apparatus for the EHFC consists of hand and foot controls monitored by a microcomputer. The assessee is required to manipulate the controls, individually and then together, to keep a red bar within a moving green box on the computer screen. The entire test takes 2½ min and a percentile score is printed automatically by the computer for the assessee (this score was used in subsequent validation analyses). The documentation provided by the test developer reports a test-retest reliability for the EHFC of .91 over a 1-month interval (VALPAR, undated). No test-retest reliability was obtained for this study because of the expense and difficulty of bringing the operators back for additional testing.

General Cognitive Ability (GCA) and Perceptual Ability (PA) (General Aptitude Test Battery) (GATB). The relevant GATB subtest raw scores were converted into normed scores (mean = 100; SD = 15) using Canadian norms developed by Hay and Barkow (1985). The subtests comprising the GCA composite consisted of the summed G, V, and N normed subtest scores from the GATB and the PA composite consisted of summed S, P, and Q normed subtest scores. The GATB composite for the purpose of this study (abbreviated as GATB in our results) was calculated by summing the GCA and PA composites into a single score. The separate composites of GCA and PA were correlated at .67 ($p < .001$), suggesting considerable redundancy of the GCA and PA composites.

Performance-Oriented Structured Interview (POSI). Ten behavioral description-type questions were written for the POSI to assess the interpersonal skills required for the job of transit operator. After asking each question and then probing for the necessary additional behavioral information for that question, the interviewer scored the interviewee on one or more dichotomously scored (i.e., 0 or 1) performance standards that were identified by supervisors and managers as typical performance required as part of the operator’s job. Fourteen standards attached to 10 interview questions were summed into a composite score assessing interpersonal skills (POSI-I). Nineteen dichotomously scored performance standards were summed into a composite score assessing adaptive skill performance (POSI-A).

A second rater listened to audio recordings of a subset of 31 interviews out of the total of 125 conducted and independently scored interviewees on the POSI-I and POSI-A scores. Correlation of the independent ratings of the main and secondary interviewers for this smaller perfor-

mance sample yielded an intercoder reliability of .88 ($p < .001$) for the POSI-I and .82 ($p < .001$) for the POSI-A.

Canadian Occupational Interest Inventory (COII). This is an inventory developed by the Canadian Employment and Immigration Commission (Booth, Begin, & Lavallee, 1981) to assess individual preferences for differing types of work activity. The COII is an ipsative measure based on the work of Cottle who identified seven bipolar factors of work preference. Booth *et al.* (1981) incorporated five of these factors into the COII, requiring respondents to choose sets of work activities that they prefer after comparing two triads of work activity representing the opposite poles of a Cottle bipolar factor. Fourteen items were constructed for each of the five bipolar factors from the Cottle model. On endorsing a triad of work activity over an opposing activity triad, a respondent is given a “1” for the triad endorsed (e.g., Business Contact) and “0” for the opposing but non-endorsed triad (e.g., Scientific). The project consultants identified three of the five COII factors as being relevant to the transit operator job based on the job analysis: Business Contact vs. Scientific; Routine vs. Creative; and Social vs. Solitary. After scoring the 14 items in each of the three subscales in the direction of Business Contact – Routine – Social, a composite COII score was computed. This composite score was then used in the validation analysis. Conceptually, this score represents the closeness of match of individual self-reported occupational interests to the content and demands of the Transit Operator job as assessed by the job analysis. Reliabilities for the COII subscales reported by Booth *et al.* (1981) are of two types: internal consistency and test-retest. Depending on the COII factor studied, the subscales used in this study reported by Booth *et al.* (1981) had internal consistency reliabilities ranging from .73 to .91 and test-retest reliabilities ranging from .75 to .91.

Criterion Variables

ISRC. Eighty-two behavioral performance standards were identified for actual operator behavior with the assistance of company supervisors and managers (e.g., “Operator sets fare sign in window appropriate to time of day [peak, off-peak] and route”) and incorporated these behavioral standards into the ISRC form. After the ISRC form was developed and piloted by the project consultants, it was administered by raters in the employ of an independent company that performs transit rating checks as its primary service line. These raters, who were either previously experienced in conducting transit ISRCs, or were novice raters conducting riding checks for the first time, were given an extensive training program by the project consultants in the proper use of the ISRC. After training, these raters then boarded the buses driven by 172 operators participating in the validation project, observed their behavior in operating the bus and dealing with

Table 1. Descriptive statistics for predictor and criterion variables

Variable	Abbreviation	<i>n</i>	Mean	<i>SD</i>	Skewness	Skewness <i>SE</i>	Kurtosis	Kurtosis <i>SE</i>
<i>Predictors</i>								
Eye-Hand-Foot Coordination	EHFC	368	41.01	36.35	.40	.13	-1.53	.25
Majority male		276	44.15	36.97	.25	.15	-1.65	.29
Women		32	29.22	33.94	1.12	.41	-.35	.81
Visible minority		47	27.45	27.70	.99	.35	-.32	.69
General Aptitude Test Battery	GATB	365	504.31	80.31	.04	.13	-.08	.26
Majority male		273	512.65	77.79	.04	.15	.23	.29
Women		32	527.26	77.55	.03	.41	-.82	.81
Visible minority		48	452.56	74.59	.27	.34	-.58	.68
Performance-oriented structured interview-interpersonal skills	POSI-I	125	8.97	2.21	-.89	.22	1.99	.43
Majority male		44	8.77	2.38	-1.28	.36	3.62	.70
Women		29	9.45	2.15	-1.35	.43	3.82	.85
Visible minority		44	9.11	1.82	-.06	.36	-.98	.70
Performance-oriented structured interview-adaptive skills	POSI-A	125	8.72	2.00	-.42	.22	.32	.43
Majority male		44	8.80	2.39	-.59	.36	.07	.70
Women		29	8.97	1.80	-.57	.43	.60	.85
Visible minority		44	8.50	1.84	-.08	.36	.28	.70
Canadian occupational interest inventory	COII	372	20.47	6.88	.07	.13	-.32	.25
Majority male		279	19.48	6.73	.10	.15	-.32	.29
Women		32	25.00	6.60	.12	.41	-.34	.81
Visible minority		49	22.76	6.81	-.07	.34	-.58	.67
<i>Criteria</i>								
In-service riding check	ISRC	172	31.95	3.56	-.15	.18	-.07	.37
Majority male		96	31.90	3.42	-.04	.25	-.59	.49
Women		28	32.82	3.50	-.03	.44	-.25	.86
Visible minority		40	31.33	4.00	-.25	.37	.66	.73
Accidents-traffic	ACC-T	367	.01	.02	7.73	.13	95.00	.25
Majority male		277	.01	.02	8.45	.15	101.89	.29
Women		32	.02	.03	1.84	.41	2.93	.81
Visible minority		49	.01	.01	1.21	.34	.41	.67
Accidents-passenger	ACC-P	367	.00	.00	8.28	.13	80.40	.25
Majority male		277	.00	.00	8.61	.15	90.08	.29
Women		32	.00	.00	5.66	.41	32.00	.81
Visible minority		49	.00	.00	NA	NA	NA	NA
Commendations	COMM	347	.02	.04	5.71	.13	47.14	.26
Majority male		259	.02	.03	4.42	.15	27.18	.30
Women		31	.05	.90	3.60	.42	15.24	.82
Visible minority		46	.01	.02	1.51	.35	1.55	.69
Complaints	COMP	354	.00	.01	4.54	.13	23.64	.26
Majority male		260	.00	.01	5.08	.15	30.60	.30
Women		31	.00	.00	5.57	.42	31.00	.82
Visible minority		46	.00	.01	2.36	.35	4.63	.69
Absences-culpable	ABS-C	355	28.54	20.93	.76	.13	-.02	.26
Majority male		260	30.62	21.38	.66	.15	-.14	.30
Women		31	21.13	19.98	1.12	.42	.86	.82
Visible minority		46	18.22	14.40	1.15	.35	.70	.69

Table 1. Continued

Variable	Abbreviation	<i>n</i>	Mean	<i>SD</i>	Skewness	Skewness		Kurtosis	
						<i>SE</i>	Kurtosis	<i>SE</i>	
Absences-non-culpable	ABS-NC	355	20.97	14.93	1.64	.13	5.31	.26	
Majority male		260	22.88	15.30	1.76	.15	5.75	.30	
Women		31	11.55	9.96	.82	.42	-.39	.82	
Visible minority		46	13.07	9.55	.52	.35	-.99	.69	
Organizational commitment	ORGCOM	80	50.55	10.30	-1.62	.27	3.50	.53	
Majority male		28	45.21	12.47	-1.36	.44	1.91	.86	
Women		18	53.94	7.32	-.69	.54	-.59	1.04	
Visible minority		30	53.80	7.23	-1.26	.43	2.02	.83	
Job involvement	JOBINV	79	35.85	9.89	.29	.27	-.24	.54	
Majority male		28	31.50	8.43	-.02	.44	-.18	.86	
Women		18	38.67	9.46	.41	.54	-.78	1.04	
Visible minority		29	39.24	9.74	.32	.43	-.30	.85	
Work involvement	WRKINV	80	20.88	6.81	.17	.27	-.64	.53	
Majority male		28	19.50	6.14	-.02	.44	-1.02	.86	
Women		18	21.00	7.13	.22	.54	-.78	1.04	
Visible minority		30	22.90	6.83	.09	.43	-.33	.83	

Note: The means for Aboriginal people were excluded because there were less than 10 data points for every variable in the sample. NA, not applicable.

passengers, and completed an ISRC form for each operator. The operator was scored a “1” if he or she met the performance standard and “0” if he or she failed to meet the standard. Because of missing data (i.e., “inadequate opportunity to observe”), only 43 of the 82 items on the ISRC form could be used and these were combined into a composite score. The internal reliability of the ISRC items comprising the composite score was .65.

Accidents Per Month. Traffic accidents per month (ACC-T) were determined by aggregating the number of traffic accidents for the individual operator since date of hire and dividing by the operator length of tenure in months. Traffic accidents refer to incidents involving collision with or damage to other vehicles or property, including collision with bicycles and pedestrians. *Passenger accidents per month* (ACC-P) were determined by aggregating the number of passenger accidents for the individual operator from date of hire and dividing by length of operator tenure in months. Passenger accidents refer to incidents happening within or outside the bus to passengers entering, riding in, or leaving the bus.

Commendations and Complaints. Commendations per month (COMM) were computed by tallying the total number of commendations received by the operator since date of hire and dividing by tenure of the operator in months. Complaints per month (COMP) were computed by counting the number of complaints recorded (e.g., from passengers and supervisors) since date of hire and dividing by operator tenure in months.

Absences. Culpable absences (ABS-C) were determined by summing the number of absences over the previous 5-year period for which the operator was “culpable” (i.e.,

no valid reasons were offered to management by the operator to explain the absence). *Nonculpable* absences (ABS-NC) were determined by summing the number of absences over the previous 5-year period for which primary responsibility for the absence was judged by management to be beyond the control of the operator.

Organizational Commitment (ORGCOM). ORGCOM was measured by nine of the 15 items contained in the Organizational Commitment Questionnaire (OCQ) developed by Mowday, Porter, and Steers (1982). A subset of the transit operator sample completed this questionnaire ($n = 80$). The internal consistency reliability of the OCQ was .89.

Involvement. Job involvement (JOBINV) was measured by the 10-item scale developed by Kanungo (1982). The internal consistency of this scale was .85. Work involvement (WORKINV) was measured by Kanungo’s (1982) six-item instrument. A subset of the transit operator sample completed these questionnaires ($n = 80$). The internal consistency of this measure was .82.

Results

Table 1 presents descriptive statistics for the predictors and criteria studied here. Also included are separate descriptive statistics by subgroup membership. Some distributional departures from normality were found for the EHFC and POSI-I predictors. The accidents and complaints data were distributed with marked departures from normality. These criterion data were left untransformed for subsequent

Table 2. Intercorrelations of predictor and criterion variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>Predictors</i>														
1. EHFC														
2. GATB	.31*** (357)													
3. POSH	-.15 (120)	.07 (121)												
4. POSI-A	-.11 (120)	.03 (121)	.29** (125)											
5. COII	-.20*** (364)	-.25*** (365)	.10 (123)	-.03 (123)										
<i>Criteria</i>														
6. ISRC	-.01 (168)	.16* (167)	.11 (99)	-.07 (99)	.10 (170)									
7. ACC-T	.11* (361)	.06 (359)	.09 (123)	.06 (123)	-.04 (366)	-.010 (169)								
8. ACC-P	.02 (361)	.02 (359)	-.12 (123)	-.06 (123)	-.14** (366)	-.00 (169)	.04 (367)							
9. COMM	.03 (340)	.04 (338)	.15 (115)	.04 (115)	.05 (345)	.03 (158)	-.00 (343)	-.01 (343)						
10. COMP	-.12* (343)	-.01 (341)	-.05 (115)	-.11 (115)	-.01 (348)	-.15 (160)	.00 (343)	.04 (343)	-.01 (347)					
11. ABS-C	.01 (344)	-.08 (341)	-.03 (115)	-.20* (115)	-.01 (348)	.02 (161)	.06 (343)	.07 (343)	-.08 (346)	.01 (353)				
12. ABS-NC	.03 (344)	-.06 (341)	-.25** (115)	-.20* (115)	-.01 (348)	.03 (161)	.04 (343)	.03 (343)	-.14** (346)	-.07 (353)	.48*** (355)			
13. ORGCOM	-.19* (78)	-.10 (77)	.26* (78)	.19~ (78)	.41*** (79)	.04 (66)	.06 (79)	.06 (79)	.06 (75)	-.14 (76)	-.36** (76)	-.16 (76)		
14. JOBINV	-.16 (77)	-.28* (76)	.14 (77)	-.05 (77)	.26* (78)	-.11 (65)	-.17 (78)	-.17 (78)	.13 (74)	.11 (75)	-.32** (75)	-.12 (75)	.59*** (79)	
15. WRKINV	-.09 (78)	-.38** (77)	.01 (78)	-.13 (78)	.21 (79)	-.30* (66)	-.11 (79)	-.24* (79)	.08 (75)	.13 (76)	-.14 (76)	-.04 (76)	.38*** (80)	.68*** (79)

Note: Sample sizes for each correlation are reported in brackets; EHFC, Eye-Hand-Foot Coordination; GATB, General Aptitude Test Battery; POSI-I, Performance-Oriented Structured Interview-Interpersonal Skills; POSI-A, Performance-Oriented Structured Interview-Adaptive Skills; COII, Canadian Occupational Interest Inventory; ISRC, In-Service Riding Check; ACC-T, Accidents-Traffic; ACC-P, Accidents-Passenger; COMM, Commendations; COMP, Complaints; ABS-C, Absences-Culpable; ABS-NC, Absences-Nonculpable; ORGCOM, Organizational Commitment; JOBINV, Job Involvement; WRKINV, Work Involvement. ~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

analyses because any transformations would make the validation results using these variables difficult to interpret.

Table 2 presents the main validation findings of the intercorrelations found among the predictor and criterion variables. It should be noted that the sample sizes vary depending on the measures being correlated. In some cases, the sample sizes are relatively small and, as such, some of the significance tests lack the power to detect statistically significant correlations.

Conclusions

As can be seen, the pattern of results linking each predictor to the various study criteria is mixed. For example, higher scores on the GATB are associated with better driving behavior (as assessed using the ISRC) and fewer customer complaints. The GATB is also predictive of lower levels of job and work involvement. Our results generally support the findings of previous research into transit operator selection in that measures of non-intellective skills (i.e., POSI and COII) have proven predictive of transit operator performance over a reasonably comprehensive domain of job performance. The positive interview results are also consistent with meta-analytic results showing generally good prediction of job outcomes by structured interviews (Wiesner & Cronshaw, 1988; Huffcutt & Arthur, 1994), as well as prediction of social and other non-intellective constructs through the use of the structured interview (Huffcutt, Conway, Roth, & Stone, 2001).

Acknowledgements

Data collected for this study were pursuant to a research contract awarded to the consulting firm of Organization

and Management Solutions, a division of the Guelph Centre for Organizational Research Inc., Guelph, Ontario, Canada (Diane Green, Executive Director).

References

- Booth, J.A., Begin, L. and Lavalley, L. (1981) *Canadian occupational interest inventory*. Toronto, Canada: Nelson Canada.
- Fine, S.A. (1989) *Functional job analysis scales: A desk aid*. Milwaukee, WI: Sidney A. Fine Associates.
- Fine, S.A. and Cronshaw, S.F. (1999) *Functional job analysis: A foundation for human resources management*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hay, T. and Barkow, B. (1985) *Revision and re-norming of the General Aptitude Test Battery (GATB)*. Toronto, Canada: Behavioural Team, A Corporation.
- Huffcutt, A.L. and Arthur, W. Jr. (1994) Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184–190.
- Huffcutt, A.I., Conway, J.M., Roth, P.L. and Stone, N.J. (2001) Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897–913.
- Human Resources Development Canada (undated). Guidelines for the Employment Equity Act and Regulations (Guidelines 4: Collection of workforce information), Ottawa, Canada: Author. Retrieved November 30, 2004 from <http://www.hrsdc.gc.ca/asp/gateway.asp?hr=/en/lp/lo/lsw/legislation/guidelines/gdln4.shtml&chs=wzp>.
- Kanungo, R.N. (1982) Measurement of job and work involvement. *Journal of Applied Psychology*, 67, 341–349.
- Mowday, R.T., Porter, L.W. and Steers, R.M. (1982) *Employee-organization linkages: The psychology of commitment, absenteeism, and turnover*. New York, NY: Academic Press.
- VALPAR (undated). *MESA statistical information report*. Tuscon, AZ: Author International Corporation.
- Wiesner, W. and Cronshaw, S.F. (1988) A meta-analytic investigation of the impact of interview format and degree structure of the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275–290.