

# Task-Based Assessment Centre Scores and their Relationships with Work Outcomes

Duncan J. R. Jackson, *The University of Seoul*

Paul Englert, *OPRA Consulting Group*

Task-based assessment centres (TBACs) have been suggested as a viable approach to evaluation in employment scenarios. Despite such suggestions, little or no empirical evidence exists on the relationship between TBAC scores and work outcomes. A sample of managers in a New Zealand service company participated in a TBAC used for organizational diagnostics and development. Results suggested a factor structure that reflected managerial roles and an uncorrected predictive validity coefficient with job performance of .42 ( $p < .01$ ). In practical terms, TBAC ratings did not discriminate on the basis of work-irrelevant variables, including age, gender, and ethnicity.

Assessment Centres (ACs) represent an approach to behavioural evaluation in the workplace that has garnered international popularity (Thornton & Krause, 2009). ACs have been used in New Zealand for evaluation purposes (Taylor, Keelty, & McDonnell, 2002) and have been appraised with regard to their psychometric properties in New Zealand contexts (Jackson, Atkins, Fletcher, & Stillman, 2005; Jackson, Barney, Stillman, & Kirkley, 2007; Jackson, Stillman, & Atkins, 2005). One of the most commonly reported psychometric findings in the AC literature is that, among AC ratings, different-dimension same-exercise correlations tend to be stronger than same-dimension different-exercise correlations (Bowler & Woehr, 2006; Lance, Lambert, Gewin, Lievens, & Conway, 2004; Sackett & Dreher, 1982). Thus, ratings in ACs tend to accentuate performance within exercises rather than performance on the basis of allegedly stable dimensions (e.g., communication skills) assessed across exercises. Such *exercise effects* have been found in New Zealand (Jackson, et al., 2007; Jackson, Stillman, et al., 2005) as well as internationally (Lievens &

Christiansen, in press), suggesting that there are conceptual problems associated with scoring a given dimension across multiple exercises.

Exercise effects have led several researchers to voice concerns about aggregating AC scores by dimensions summarized across exercises. Particular concerns include the meaning given to dimensions in developmental ACs (Kudisch, Ladd, & Dobbins, 1997) and limitations in terms of fostering an understanding as to the mechanism underlying AC functionality (Klimoski & Brickner, 1987). In response to the measurement issues associated with ACs, three perspectives on the meaning underlying AC ratings have emerged (Lievens & Christiansen, in press). Firstly, the traditional dimension-based approach, where dimensions are thought to form meaningful constructs when aggregated across different exercises (Arthur, Day, & Woehr, 2008). Secondly, the task-based approach, which proposes that ratings from ACs should be aggregated within exercises to form meaningful exercise-based constructs (Jackson, Stillman, et al., 2005). Thirdly, the mixed-model approach, which incorporates

aspects of both dimension scores in combination with exercise scores (Hoffman, Melchers, Blair, Kleinmann, & Ladd, in press).

Despite internal measurement challenges, ACs are often found to be predictive of work outcomes. With a job performance criterion, the most recent meta-analysis suggested a criterion-related validity estimate of .36 (Arthur, Day, McNelly, & Edens, 2003) and a previous meta-analysis returned similar results with an estimate in the order of .37 (Gaugler, Rosenthal, Thornton, & Bentson, 1987). Arthur, et al. stated that their meta-analysis of the criterion-related validity of ACs focused on dimensions because of their importance to psychology and their historical use in this context (see p. 128). It is, indeed, difficult to contest the importance of the assessment of constructs such as dimensions in psychology, given the research database that has accumulated on the measurement of psychological variables. Nonetheless, some researchers have suggested that the actual constructs measured by ACs might be different from those formalized in dimension scores (Jackson, et al., 2007; Jackson, Stillman, et al., 2005; Lance, 2008a; Lowry, 1997).

The fact that dimension scores are summarized across exercises in dimension-based ACs has led to a belief that ACs tap relatively stable and enduring variables, akin to traits (Jackson, et al., 2007; Jackson, Stillman, et al., 2005; Sackett & Dreher, 1982). The concept of sampling behaviour across different measures in this way is

reminiscent of trait theories presented in keystone papers on psychometrics such as that by Campbell and Fiske (1959). The urban legend that has since emerged is that, in a multitrait-multimethod matrix tradition, AC exercises (cf. methods) are seen as mere vehicles for tapping into dimensions (cf. traits, see Lance, Baranik, Lau, & Scharlau, 2009). Under this view, dimension scores represent the ultimate outcome from an AC and exercises are seen as sources of error that potentially interfere with dimension assessment. Jackson et al. (2007) suggest that, because they are summarized across exercises, dimensions are, in practical terms, often presumed to behave like stable traits. Further, Howard (2008) suggests that ACs were never intended as trait measures, which raises questions about the appropriateness of scoring ACs by exclusively aggregating dimension scores across exercises.

Alternative directions for scoring ACs have been explored and one of these approaches focuses on aggregating scores by simulation exercise. Lance et al. (2004), Lowry (1997), and Robertson, Gratton, and Sharpley (1987), among others, have suggested that ACs could be set up in such a manner that they resemble collections of work simulations with each simulation representing a work role. Lowry (1997) provided guidelines for how such *task-specific* ACs could be set up and what their potential benefits could be. More recently, task-specific ACs have been referred to as *task-based* ACs (TBACs, Lance, 2008a; Lance, in press).

TBACs emphasize a job-related focus for ACs with the aim of presenting an internally construct valid approach that also meaningfully predicts such work outcomes as job performance (Jackson, et al., 2007; Jackson, Stillman, et al., 2005; Lance, 2008a; Lowry, 1997). Because a key aim for TBACs is to be internally construct valid, another fundamental feature is to present developmental information in a manner that is meaningful (Jackson, Stillman, & Englert, 2010). Admissible evidence for construct validity, criterion-related validity, and meaningful bases for employee development highlight issues concerning ethics in evaluation via ACs (Brannick, 2008; Jackson, et al.,

2007). Valid and job-relevant predictors are needed to maintain an ethical approach to selection and development in organizations. This is especially important in a multicultural setting like that in New Zealand where cognitive ability scores have been found to vary among ethnic groups. Such variability may be associated with adverse impact (Guenole, Englert, & Taylor, 2003). Where evidence is found for adverse impact, TBACs provide an alternative form of assessment with the potential to avoid such biases (Jackson, et al., 2007; Lance, 2008a; Lowry, 1997). However, these potential features, some of which are merely claimed, need to be verified empirically.

Proponents of the TBAC approach describe a theoretical outlook on the AC technique that differs from traditional approaches. Under a TBAC guise, ACs are regarded as collections of work simulations that tap work-related roles (Jackson, in press). The unit of measurement here is an interaction between people and different situations. Here, variability associated with different situations (exercises) is regarded as acceptable (see Jackson, et al., 2005). The output for a TBAC is an overall score for each exercise that represents one or more work role and an overall rating based on multiple exercise scores (Jackson, in press; Lance, Foster, Nemeth, Gentry, & Drollinger, 2007). Frameworks for prospective work roles measured in TBACs are available in the research database, and include those famously suggested by Mintzberg (1971, 1973).

Mintzberg (1971, 1973) describes three-meta roles, each of which is reflective of a set of sub-roles, including information-processing roles (disseminator, monitor, spokesperson), decision-making roles (entrepreneur, disturbance handler, resource allocator, negotiator), and interpersonal roles (liaison, figurehead, leader). Mintzberg's roles have enjoyed wide application in the management literature (Pearson & Chatterjee, 2003) and, in addition, have been applied to commonly used AC exercises (Shapira & Dunbar, 1980). Other, analogous frameworks are also available for the researcher, including the nine managerial position duties and responsibilities (supervising, planning

and organizing, decision-making, monitoring indicators, controlling, representing, co-ordinating, consulting, and administering) detailed in Yukl (2010, p. 84). Joyce et al. (1994), Russell and Domm (1995) and Hogan, Broach, and Salas (1990) also provide frameworks that could also be applied to TBAC scores.

With regard to criterion-related validity, the foundation for a TBAC approach stems from *behavioural consistency* (see Jackson, in press; Wernimont & Campbell, 1968). This is the same theoretical justification for behavioural techniques such as work samples and situational exercises, where performance in a work-relevant simulation is likely to predict job performance in similar situations. Thus, it is the interaction between the person and the situation that is the behavioural predictor under this approach. Validities for work samples are often among the highest found in predictors in employee selection. Schmidt and Hunter (1998) report that work sample tests yielded the highest predictive validities available across a range of selection devices with an  $r$  of .54 with a job performance criterion (note, however, that Roth, Bobko, & Mcfarland, 2005 provide a more conservative estimate of .33). Work samples also have the advantage of being associated with lower levels of adverse impact than many other available predictors (Robertson & Kandola, 1982). In line with behavioural consistency, Sackett and Harris (1988) suggested focusing on situational responses in ACs to help resolve the exercise effect issue. In a similar vein, Lance and colleagues (Lance, 2008a, 2008b; Lance, et al., 2004; Lance et al., 2000) have suggested that further research into a TBAC approach is warranted.

A potential challenge for the TBAC conceptualization is a lack of empirical research justifying it as a useful tool for employment decisions (Lance, in press). Lance et al. (2004, p. 383) states that, over and above Lowry's (1997), primarily anecdotal, findings about TBACs, there is "little or no additional evidence on their reliability and validity". Some of the available research on the psychometric properties of this approach was published in Jackson,

Stillman, et al. (2005). As part of their study, the authors presented favourable factor analytic and Generalizability-theory-based evidence. However, no information was presented on work-related outcomes such as adverse impact or on criterion-related validity. As such, the TBAC approach to AC design could still be regarded as uncharted territory for practical purposes.

Given the exercise effect issues found in dimension-based ACs, investigations into available alternative approaches appear to be justified. The present study seeks to examine both the internal psychometric properties of a TBAC and, importantly, the work-related outcomes associated with the approach. Of interest here, also, is whether exercise-driven scores can be interpreted as meaningful constructs (Arthur & Villado, 2008). The idea that exercise scores might represent work roles has yet to be formally tested. In this respect, Russell and Domm (1995) found evidence for role constructs. However, they used *post-consensus dimension ratings*, which means that aggregate role scores were generated at the end of the AC rather than during each exercise. Under a post-consensus approach, it is impossible to separate dimension and exercise effects and, as such, Russell and Domm were unable to empirically test whether their roles were based on exercises or on some other foundation.

Evidence for exercise-based role constructs would be gained if exercise scores map, in theoretical terms, onto the role frameworks presented in the extant literature (e.g., Mintzberg, 1971, 1973; Yukl, 2010). In the present study, we opted to focus on Mintzberg's broad role framework because of its wide application to managerial scenarios (Pearson & Chatterjee, 2003) and to AC exercises (Shapira & Dunbar, 1980). Expanding upon previous findings showing exercise-based structures for AC data (Jackson, et al., 2007; Jackson, Stillman, et al., 2005; Joyce, et al., 1994; Lance, et al., 2007; Lance, et al., 2004; Lance, et al., 2000; Robertson, et al., 1987; Sackett & Dreher, 1982; Stillman & Jackson, 2005), we hypothesise the following:

*Hypothesis 1.* Ratings from a TBAC will reflect scores on exercises

that can be interpreted as representing aspects of Mintzberg's (1971, 1973) managerial role framework.

The work sample literature bears many similarities to ideas presented in the TBAC approach. In particular, the notion of behavioural consistency seen in work samples is also found in TBACs (Jackson, in press). Currently, little or no published evidence exists for the criterion-related validity of TBACs. Moreover, little or no evidence exists for or against the degree of adverse impact associated with TBACs. This has important implications for New Zealand, given the diverse ethnic groups resident in the country. In fact, Guenole, Englert, and Taylor (2003) found differences among ethnic groups with regard to their cognitive ability scores in a New Zealand sample. Encouragingly, evidence exists for the criterion-related validity of work samples as well as for their capacity to avoid adverse impact when compared to other forms of psychological assessment (Robertson & Kandola, 1982; Roth, et al., 2005; Schmidt & Hunter, 1998). Given the conceptual links between TBACs and work samples, we hypothesize the following:

*Hypothesis 2.* That TBAC scores will show evidence of a meaningful relationship with scores on a job performance criterion.

From an exploratory perspective and given the links between TBACs and work samples mentioned above, we also aimed to investigate the following research question:

*Research Question 1.* Are TBAC scores associated with patterns of variation that might suggest minimal adverse impact (including on work-irrelevant criteria such as gender, age, and ethnicity)?

## Method

### Participants

A total pool of 229 managers from a large nation-wide organization in New Zealand participated in an AC used for a range of employment decision-making purposes, including the assessment of organizational strategy and developmental needs. The

organization under study specialized in insurance, credit, banking, postal, and administrative services. Of the total number of participants, 214 were retained for data analysis, the remainder being too incomplete for inclusion. The mean age of participants was 45.53 ( $SD = 10.33$ ), and around 54% of the sample were male and 46% were female. Different ethnicities in the sample included New Zealand European (71%), Maori and Pacific Island (9%), European (7%), Indian (9%), and Other (4%). Just over half of the participants had completed high school (around 53%) or held a trade certificate or degree (around 25%).

### Measures

*Job analysis.* Development of the AC was guided by Lowry's (1997) course of action for TBACs. Each exercise in a TBAC is treated as a substantive measure in a similar manner to a collection of situational exercises. Exercise development primarily involved the use of subject matter expert interviews and task analyses which included teams of human resource managers, area managers, and line managers. Williams and Crafts' (1997) framework for conducting inductive job analyses was used to guide this process.

*Assessment exercises.* Four exercises were developed in total. Each exercise carried an associated 10-item behavioural checklist where performance was rated from 1 (certainly below standard) to 6 (certainly above standard). Examples of checklist items included "Uses objective and non-emotive language when delivering feedback to others" and "Comes up with solutions that have the customer in mind". Dimension titles were also developed that were associated with each behavioural descriptor in the AC. However, the focus of the present study was on the TBAC component of the AC rather than on dimension titles (the details of this focus are discussed later). Details of the exercises are presented below.

*Exercise 1: Managing new staff.* The first part of this exercise involved a discussion in small groups on managing new front-line staff members. The second part of this exercise involved an individual presentation to assessors

on important factors to consider when managing new staff.

*Exercise 2: Selecting new staff.* The format of this exercise was the same as that above and included a discussion and individual presentation. The focus here, however, was on factors that participants would consider important for the selection of new staff specializing in insurance and lending.

*Exercise 3: Photo exercise.* The photo exercise constituted a group discussion and debate in which a series of photos were shown to participants displaying the interior and exterior of retail stores. Access problems and issues around aesthetics were purposely staged in the photos to provide material for debate.

*Exercise 4: Coaching exercise.* The coaching exercise was a role play in which candidates were required to plan a performance coaching meeting with an employee. Participants then were asked to role play a coaching session in which performance plans for the next six months were to be agreed.

*Job performance.* Job performance ratings were taken from the organization's existing performance management system. This involved a behaviourally anchored rating scale (see Borman, 1986) ranging from 1 to 150 with behavioural descriptors at intervals detailing actions that were associated with particular ranges of scores. For example, higher scorers were described as having "contributed more widely than is expected of their role/level" and "over-achieved on their targets by a significant margin". The potential for criterion contamination, i.e., knowledge of AC ratings influencing job performance ratings (see Klimoski & Brickner, 1987), was unlikely in this study because AC ratings were used for specific developmental requirements only and were kept separate from job performance ratings. Moreover, in the predictive validation reported later, an entire year had passed before job performance ratings were completed.

*Assessors.* Assessors either ranked a management level above participants ( $n = 19$ ) or were consultant psychologists ( $n = 4$ ). The assessor-to-participant ratio was 1:2 (consistent with the International Task Force on Assessment Center Guidelines, 2009) and assessors

were rotated to help minimize the effects of rater-specific error. Assessor training involved the use of an adapted frame-of-reference training procedure (see Lievens, 1998; Pulakos, 1986; Schleicher, Day, Mayes, & Riggio, 1999). Training lasted for a two-day period and covered familiarization with rating instruments and exercise format, common rater errors, and, importantly, practice assessments involving mock candidates. Practice assessments covered all exercises and provided rating data for standardization discussions among assessors. Executives from the participating organization were in attendance during these discussions to help ensure that schemata around performance norms were in line with organizational expectations.

### Data Analysis

An exploratory factor analysis (EFA) was used, initially, in order to ascertain the overall structure of the behavioural checklist ratings, from a data-driven perspective. In turn, confirmatory factor analysis (CFA) was used to provide structural evidence, in line with Brown (2006), who suggests applying CFA after gaining evidence from the EFA perspective. The goodness-of-fit indices we included and their guideline cut-off criteria (see Hu & Bentler, 1998, 1999) were: the standardized root mean square residual (SRMSR,  $\square .08$ ), the root mean square error of approximation (RMSEA,  $\square .06$ ), the Tucker-Lewis Index (TLI,  $\geq .95$ ), the comparative fit index (CFI,  $\geq .95$ ), and the Akaike information criterion for comparisons among non-nested models (AIC, where relatively smaller coefficients indicate better fit). To assess criterion-related validity, bivariate correlations were computed along with corrections for criterion unreliability and range restriction where relevant. Also, where appropriate, a multivariate perspective was indicated. To assess indicators that might be associated with adverse impact, parametric and nonparametric mean comparison tests were utilized where appropriate. CFAs were performed using AMOS (version 18). All other analyses were performed using PASW (version 18).

## Results

### Data Structure

In the interests of maintaining a data-driven approach, an EFA (principal axis factoring with a latent root criterion, see Spicer, 2005) was initially applied to the AC ratings. Direct oblimin (i.e., oblique) rotation was used, given the correlated exercise effects reported previously by Lance, Noble, and Scullen (2002). Table 1 shows the factor structure of the TBAC ratings, which mostly represented clean factors with behavioural items nested within their respective exercises. Note, as an aside, that this analysis met the criteria for using EFA with small sample sizes (as set out in de Winter, Dodou, & Wieringa, 2009).

From this point, we tested three complementary CFA models. To allow for reasonable subject-to-variable ratios, three item parcels were entered as observed variables for each latent exercise factor (see Bandalos & Finney, 2001). This process involved the use of a random number table to allocate items into parcels. The first CFA was analogous to a null comparative model and consisted of no exercise (0E) factors and only one overall performance dimension (1D, i.e., a 0E1D model). As expected, this yielded unacceptable fit estimates ( $\chi^2(54) = 1119.08, p < .05$ ; SRMSR = .157, RMSEA = .304, TLI = .397, CFI = .507, AIC = 1167.080). The second CFA most closely replicated the data-driven EFA findings and included four correlated exercise factors (4E0D). This yielded an acceptable fit ( $\chi^2(48) = 61.59, ns$ ; SRMSR = .031, RMSEA = .036, TLI = .991, CFI = .994, AIC = 121.594). Correlations among latent exercise factors ranged from .34 to .58 ( $M_r = .48, SD_r = .09$ ). On the subject of relationships among latent exercise factors, Lance et al. (2007) found evidence for a general performance dimension in their exercise-based AC scores. As such, our third CFA also tested for general performance by augmenting the 4E0D model with a single, first order, overall performance dimension. The resulting 4E1D model yielded an acceptable fit that was marginally better than that for the 4E0D model ( $\chi^2(36) = 36.82, ns$ ; SRMSR = .020, RMSEA = .010, TLI = .999, CFI = 1.000, AIC = 120.819). A  $\chi^2$  difference

Table 1. Exploratory Factor Analysis of Task-Based Assessment Center Ratings

Item	Factor					$h^2$	$M$	$SD$
	1	2	3	4	5			
Ex1b1				.79		.69	3.33	1.29
Ex1b2				.62		.67	3.74	1.28
Ex1b3				.70		.55	3.37	1.10
Ex1b4				.66		.62	3.29	1.23
Ex1b5				.64		.43	3.34	1.23
Ex1b6				.63		.40	3.01	1.21
Ex1b7				.47		.34	3.89	1.08
Ex1b8				.51		.64	3.32	1.35
Ex1b9				.76		.71	3.34	1.21
Ex1b10				.79		.74	3.09	1.29
Ex2b1			.81			.67	3.32	1.19
Ex2b2			.59			.68	3.29	1.41
Ex2b3			.38		.35	.35	3.89	1.04
Ex2b4			.57			.55	3.49	1.29
Ex2b5			.61			.37	3.45	1.12
Ex2b6			.81			.63	3.27	1.11
Ex2b7			.83			.74	3.23	1.24
Ex2b8			.70			.67	3.34	1.24
Ex2b9			.84			.72	3.35	1.21
Ex2b10			.67			.68	3.70	1.24
Ex3b1		.80				.66	2.91	1.27
Ex3b2		.72				.64	2.79	1.24
Ex3b3		.79				.64	2.88	1.20
Ex3b4		.77				.60	3.10	1.14
Ex3b5		.40				.23	3.69	1.07
Ex3b6		.75				.63	3.26	1.31
Ex3b7		.74				.64	3.15	1.44
Ex3b8		.82				.66	2.88	1.15
Ex3b9		.85				.78	2.77	1.20
Ex3b10		.79				.70	3.27	1.38
Ex4b1	.79					.71	3.10	1.32
Ex4b2	.74					.66	3.21	1.24
Ex4b3	.68					.49	3.43	1.32
Ex4b4	.86					.70	2.88	1.38
Ex4b5	.73					.62	2.86	1.36
Ex4b6	.76					.60	3.51	1.19
Ex4b7	.75					.58	3.22	1.31
Ex4b8	.80					.72	2.87	1.28
Ex4b9	.84					.72	2.97	1.22
Ex4b10	.72					.55	3.50	1.21
SS	9.24	8.78	8.44	9.13	1.42			
%	35.37	45.95	53.82	59.48	60.90			

Principal axis factoring, oblique rotation, latent root criterion.  $h^2$  = communality, SS = sums of squared loadings, % = cumulative variance explained. Ex1 to Ex4 = exercises, b1 to b10 = behavioural items. Loadings < .3 were suppressed for clarity.

test between the 4E0D model and the 4E1D model was statistically significant ( $\chi^2(12) = 24.77, p < .02$ ) providing further evidence that the 4E1D model fit observed data better than the 4E0D model.

Coefficient alpha was used to estimate the internal consistency of each exercise factor ( $\alpha = .92, .92, .93, \text{ and } .94$  for exercises 1 through 4, respectively). The internal consistency of the overall exercise rating (described below) was estimated using the average alpha across exercises ( $\alpha = .93$ ). These estimates met criteria for admissibility associated with coefficient alpha (Lance, Butts, & Michels, 2006). The reliability of job performance was estimated by the internal consistency between the single-item job performance measure ( $M = 104.03, SD = 15.82$ ) and a second, more general job performance item on a 4-point scale that participants were also rated on ( $M = 2.93, SD = .36$ ). The standardized coefficient alpha (see Kline, 1999) here was considered acceptable ( $\alpha = .84$ ).

In concert, the evidence above can be taken to suggest that the exercise-based scores in this study could be considered as potentially meaningful constructs. Drawing from Mintzberg's (1971, 1973) role framework, Exercise 1 centred on managing new staff. As such, the pattern of behaviours in this exercise factor can be interpreted as aligning conceptually to components of the *leader* role, where managers are required to provide motivation and favourable conditions for subordinates. Exercise 2 covered selecting new staff, which aligns to parts of the *resource allocator* role, whereby managers need to make decisions about how to optimally utilise human resources. This also comprises another part of the *leading* role, particularly the focus on hiring. The main topic of Exercise 3 was about identifying problems in outlet stores and, as such, aligns most closely with aspects of the *disturbance handler* role. This involves identifying problems and managing them to the benefit of the organization. Exercise 4 centred on coaching and setting performance plans for existing staff. These activities also align with characteristics of yet another part of the *leading* role, particularly the component that deals with providing

guidance to employees.

In sum, the ratings in this TBAC appeared to show reasonable structural evidence and reflected a relatively clean set of correlated exercise scores from both data-driven and a priori modelling perspectives. These exercise factors were readily interpretable as reflecting aspects of management roles from the Mintzberg framework (1971, 1973). Evidence was also found for an overall performance dimension, reflected in the correlated nature of the exercise scores and formally tested in the 4E1D CFA model. Taken together, these findings present evidence largely in favour of Hypothesis 1.

### Criterion-Related Validity

The relationship between AC ratings and job performance ratings taken a year later was investigated using correlational approaches. For the criterion-related validity study, a total of 100 participants were found to have matching TBAC and job performance scores. This was due to natural attrition over the course of a year and to a number of company branches that did not use the performance management system described here. Because the 4E1D CFA model in the CFA analyses above returned an acceptable fit, it was deemed acceptable to create an overall exercise rating (OER, i.e., the average exercise score). The uncorrected predictive validity coefficient between the OER and job performance was  $r = .42 (p < .001)$ . When corrected for range restriction and criterion unreliability, this coefficient rose to  $r = .52$  (see Gatewood, Feild, & Barrick, 2008; Muchinsky, 1996). Note that the OER SDs for the unrestricted and restricted groups were not vastly different in practical terms ( $SD_u = 0.74, SD_r = 0.68$ ) and criterion reliability was estimated by the standardized alpha computed for OJP (standardized alpha = .84). To assuage concerns about data distributions, we also ran a nonparametric correlation (Siegel & Castellan, 1988) between the OER and the job performance criterion. This was similar to the Pearson correlation presented above (uncorrected Spearman's  $\rho = .40, p < .001$ ).

Using structural equation modeling, we also tested predictive validity by augmenting the 4E1D model with the

job performance measure. Measurement error was incorporated into this structural model by setting the unstandardized error of job performance to the variance (of job performance) times 1 minus the internal consistency estimate of .84 (see Bollen, 1989; Brown, 2006). The results of this analysis suggested that the OER explained around 52% of the variance in job performance (standardized  $\beta = .72, p < .05$ ). The solution here was admissible and model fit statistics were within acceptable ranges ( $\chi^2(47) = 39.57, ns$ ; SRMSR = .032, RMSEA =  $< .001$ , TLI = 1.013, CFI = 1.00).

Within-exercise aggregates were also regressed on the job performance variable. From the multiple regression viewpoint, the four exercises collectively returned an adjusted  $R^2$  of .16 (a multiple  $R$  of .44,  $p < .001$ ). Standard beta weights among exercises were as follows;  $\beta$  (exercise 1) = .18;  $\beta$  (exercise 2) = .14;  $\beta$  (exercise 3) = .20;  $\beta$  (exercise 4) = .05. Independently, none of these betas was statistically significant. Almost identical results were arrived at through a matching structural model (the associated conclusions albeit tentative because of limited subject-to-variable ratios). This suggests that only the OER was justifiable for making decisions around the prediction of job performance (e.g., selection or promotion decisions).

A small concurrent dataset was also available, including 29 participants who had their job performance assessed around the same time as when the AC took place. As a supplementary examination, a non-parametric correlation analysis on this dataset revealed a very similar criterion-related effect size to that presented above for the predictive validation (Spearman's  $\rho = .42, p < .05$ ). Note that this correlation was not corrected for range restriction or unreliability in the criterion. Taken together, the predictive and supplementary concurrent validation evidence presents, at least, initial evidence for Hypothesis 2.

### Indicators of Adverse Impact

Patterns of variation that might act as indicators of adverse impact were investigated by analyzing demographic differences in OER scores. A Kruskal-Wallis test revealed no significant differences in OER scores across the

various ethnicities included in this study ( $\chi^2(4) = 5.72, ns$ ). Frequencies of OER scores from males ( $n = 115$ ) and females ( $n = 99$ ) were fairly well balanced  $N$ -wise. A  $t$  test revealed no significant difference in scores across genders ( $t = -1.24, df = 211.48, ns$ ). A statistically significant negative correlation was observed between age and OER scores ( $r = -.18, p < .05$ ), however the attenuated effect size here suggested it was unlikely that the AC discriminated on the basis of age in practical terms. With regard to magnitude, the absence of any notable relationship between TBAC ratings and work-irrelevant criteria provides initial and exploratory support for the notion that TBACs scores minimise effects that could be associated with adverse impact (see Research Question 1).

## Discussion

The debate on ACs has resulted in the emergence of three major perspectives on the meaning underlying AC ratings (Hoffman, et al., in press; Jackson, in press; Lance, 2008a). Firstly, the dimension-based perspective posits that ACs ratings reflect dimensions that represent meaningful scores when aggregated across different exercises. Secondly, the TBAC approach posits that AC ratings are best considered as being specific to particular exercises. Under this approach, exercises are considered to be substantive measures of work roles. Thirdly, the mixed-model AC approach posits that AC ratings are best represented by considering both dimensions and exercises together. Our contention is that these positions should not be considered as three distinct categories. Rather, they appear to be anchors on a continuum with the dimension and task-based perspectives at the extremes and the mixed-model approach at the centre.

The results of the present study provide psychometric evidence for a class of AC that errs towards a TBAC perspective but the results also indicate the addition of a dimension-based element. In support of Hypothesis 1, results suggested that from both data-driven (EFA) and a priori modelling (CFA) perspectives, a structure based on exercises yielded an admissible fit for the dataset. It was possible to interpret the exercise factors as representing

aspects of the role framework presented by Mintzberg (1971, 1973), across Exercise 1 (*Leader Role* – motivating and providing favourable work conditions), Exercise 2 (*Resource Allocator Role* – optimally utilising human resources; *Leader Role* – hiring employees), Exercise 3 (*Disturbance Handler Role* – identifying and managing problems in the organization), and Exercise 4 (*Leader Role* – guiding employees). Focusing on the covariance modelling perspective, exercise scores were found to be intercorrelated ( $M_r = .48, SD_r = .09$ ), as has been observed in previous studies, suggesting the existence of a general performance factor (Lance, et al., 2007). A covariance structure that reflected the four exercises in this AC plus a general performance factor (4E1D) fit observed data reasonably well (for 4E1D,  $\chi^2(36) = 36.82, ns$ ; SRMSR = .020, RMSEA = .010, TLI = .999, CFI = 1.000). Using the AIC as an index for model comparison, the 4E1D model (AIC = 120.819) fit observed data considerably better than the 0E1D model (AIC = 1167.080) and marginally better than the 4E0D model (AIC = 121.594). Although, a  $\chi^2$  difference test indicated that the 4E1D model fit observed data significantly better than the 4E0D model ( $\chi^2(12) = 24.77, p < .02$ ). These results support a model that leans towards a TBAC view, but also incorporates a general performance dimension. As such, the results support a TBAC-oriented mixed-model stance.

With regard to Hypothesis 2, evidence was also garnered for work outcomes associated with the 4E1D model, with an uncorrected predictive validity coefficient of .42 with a job performance criterion (.52 when corrected for range restriction and unreliability in the criterion measure). A supplementary (and small) concurrent validation study yielded a similar result (uncorrected Spearman's  $\rho = .42, p < .05$ ). Further highlighting the importance of a mixed-model perspective on these data, multiple regression analyses revealed that a linear composite based on all four exercises was a significant predictor of job performance. Individual exercises were, however, not significant predictors of this criterion.

With regard to potential indicators of adverse impact and Research Question

1, TBAC ratings were not significantly associated with discrimination on the basis of gender and ethnicity. The issue of ethnicity, in particular, as it relates to psychometric testing, presents an important topic for New Zealand employment scenarios. Differences in cognitive ability scores have been found across different ethnicities in a New Zealand sample (Guenole, et al., 2003). Perhaps, as with findings commonly reported in the literature on work samples (Robertson & Kandola, 1982; Roth, et al., 2005), TBAC-oriented approaches present an alternative form of assessment that have the potential to alleviate problems associated with adverse impact. A significant negative correlation was found between age and OERs ( $r = -.18, p < .05$ ). Nonetheless, this effect size only constituted around 3% of variance explained and, as such, was probably not of any practical substance.

Task, dimension, and mixed-model AC approaches have been presented as distinct categories in the past. The results of the present study suggest that these interpretations of AC-derived ratings differ only by relative emphasis. The AC literature, in our view, needs to amalgamate in order understand the features of the different emphases and under what conditions they might be more or less appropriate. The task-based emphasis is probably one of the least-researched approaches to AC design and additional insights into TBACs are sorely needed (Jackson, et al., 2007; Lance, 2008a). While some structural evidence for TBACs exists (Heo & Shin, 2010; Jackson, Stillman, et al., 2005; Lance, et al., 2007) we do not know of any previous studies that have investigated work-related outcomes associated with this approach.

The paucity of available TBAC research on work outcomes is one possible reason that they have not yet been taken up into mainstream AC practice. Another possible issue that might act as a barrier to the use of TBACs includes a potential lack of concordance between popular competency models that underlie assessment in many human resource management scenarios and task-based ratings (Hoffmann, 1999; Jackson, et al., 2007; Markus, Cooper-Thomas, & Allpress, 2005). The

interpretation of latent exercise factors as work role measures in this study presents a potential bridge between competencies and TBACs. There are a number of work role frameworks available and there is the potential for conceptual links between these and available competency frameworks (Hogan, et al., 1990; Joyce, et al., 1994; Mintzberg, 1971, 1973; Russell & Domm, 1995; Tett, Guterman, Bleier, & Murphy, 2000; Yukl, 2010).

### Limitations and Future Directions

The results presented here emanate from a single validation study on one organization. As such, this raises questions about cross-sample generalization and whether similar results would be obtained in different industries. There are two key responses to this as a potential criticism. Firstly, there are presently few or no studies on TBACs incorporating information on work outcomes (Jackson, et al., 2007; Jackson, Stillman, et al., 2005; Lance, 2008a). As such, the results here present, in the least, an encouraging start point for further evidence in other contexts. Secondly, the participant organization and the individual participants in this study were involved in a range of different service activities. Although other research on work outcomes associated with TBACs is necessary, the participants in this study did represent employees from a range of work activities. Also, on the subject of cross-sample generalization, other measures of job performance and additional indicators of adverse impact could be integrated into future studies in order to expand upon the findings here.

Another potential limitation associated with our results stems from the idea that the AC in our study was not a purist TBAC. To elaborate, the AC here was designed in accordance with existing guidelines on the development of TBACs (Lowry, 1997) but included dimension titles that were associated with TBAC ratings. While this is true, we argue that, in applied settings, there has never been and possibly never will be a *pure* TBAC. Likewise, we also argue that there will possibly never be a *pure* dimension-based AC either. Almost all empirical studies on AC ratings have found factors that resemble *correlated* exercise effects (implying

the presence of an underlying general performance dimension, see Lance, 2008a; Lance, et al., 2007; Lance, et al., 2004; Sackett & Dreher, 1982; Sackett & Lievens, 2008) or a mixture of exercise and dimension factors (Bowler & Woehr, 2006; Hoffman, et al., in press). Given the task-driven approach to the design and implementation of this AC, coupled with empirical evidence for an exercise-based structure, the relative emphasis here was deemed to incline towards the TBAC approach. Yet, this was not a pure TBAC and is better described as a mixed-model AC with a task emphasis. In the light of the results of this study and new AC research (Hoffman, et al., in press), we doubt that a purist TBAC approach is a realisable position anyway and, rather, it probably represents an academic extreme at one end of a continuum of AC design types (as described earlier).

TBACs present somewhat of a green field in terms of research opportunities. Potential studies include extensions and cross-sample generalizations of the results found here. In addition, it would be interesting to learn more about the cognitive processes, from the perspective of assessors, involved in generating ratings in TBACs. Further information is also required on training for TBACs and whether they are, as some commentators claim, easier to train assessors on than ACs that have more of a dimension focus (Lance, 2008a; Lowry, 1997). Additional research is also required on role taxonomies and whether it would be appropriate to integrate existing taxonomies into one that is purpose-built for TBACs. From a practical perspective, it would also be interesting to learn more about how such task taxonomies can form links with competencies in a manner that moves beyond theoretical links and applies an empirical basis (cf. Markus, et al., 2005; Ruth, 2006).

### Conclusion

A task-based approach to ACs has been suggested in theoretical terms for over two decades (Goodge, 1988). Yet, prior to the present study, little or no empirical evidence of the relationship between TBAC scores and work outcomes has been published. Here, evidence was found for a notable

correlation between TBAC scores and job performance and TBAC scores were not found to be notably related to indicators of adverse impact. Overall, the present results show encouraging psychometric evidence for a TBAC approach. These results bear relevance to multi-cultural contexts like New Zealand where indicators of adverse impact in psychometric testing has been found to present potential ethical challenges (Guenole, et al., 2003).

### References

- Arthur, W., Jr, Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.
- Arthur, W., Jr, Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternative view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 105-111.
- Arthur, W., Jr, & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435-442.
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Shumaker (Eds.), *Advanced structural equation modeling: New developments and techniques* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Borman, W. C. (1986). Behavior-based rating scales. In R. A. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 100-120). Baltimore, MD: Johns Hopkins University Press.
- Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology, 91*, 1114-1124.
- Brannick, M. T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 131-133.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.



- de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*, 147-187.
- Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2008). *Human resource selection* (6th ed.). Mason, OH: Thomson South-Western.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493-511.
- Goodge, P. (1988). Task-based assessment. *Journal of European Industrial Training, 12*, 22-27.
- Guenole, N., Englert, P., & Taylor, P. J. (2003). Ethnic group differences in cognitive ability test scores within a New Zealand applicant sample. *New Zealand Journal of Psychology, 32*, 49-54.
- Heo, C. G., & Shin, K. H. (2010, June). *Reliability and validity of nested-designed assessment center*. Paper presented at the Korean Society for Industrial Organizational Psychology, Daejeon, Korea.
- Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. T. (in press). Exercises and dimensions are the currency of assessment Centres. *Personnel Psychology*.
- Hoffmann, T. (1999). The meanings of competency. *Journal of European Industrial Training, 23*, 275-285.
- Hogan, J., Broach, D., & Salas, E. (1990). Development of a task information taxonomy for human performance systems. *Military Psychology, 2*, 1-19.
- Howard, A. (2008). Making assessment Centres work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 98-104.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. *Structural Equation Modeling, 6*, 1-55.
- International Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment, 17*, 243-253.
- Jackson, D. J. R. (in press). Theoretical perspectives on task-based assessment Centres. In D. J. R. Jackson, C. E. Lance & B. J. Hoffman (Eds.), *The psychology of assessment centers*. New York: Routledge.
- Jackson, D. J. R., Atkins, S. G., Fletcher, R. B., & Stillman, J. A. (2005). Frame of reference training for assessment Centres: Effects on interrater reliability when rating behaviors and ability traits. *Public Personnel Management, 34*, 17-30.
- Jackson, D. J. R., Barney, A. R., Stillman, J. A., & Kirkley, W. (2007). When traits are behaviors: The relationship between behavioral responses and trait-based overall assessment center ratings. *Human Performance, 20*, 415-432.
- Jackson, D. J. R., Stillman, J. A., & Atkins, S. G. (2005). Rating tasks versus dimensions in assessment Centres: A psychometric comparison. *Human Performance, 18*, 213-241.
- Jackson, D. J. R., Stillman, J. A., & Englert, P. (2010). Task-based assessment Centres: Empirical support for a systems model. *International Journal of Selection and Assessment, 18*, 141-154.
- Joyce, L. W., Thayer, P. W., & Pond, S. B. (1994). Managerial functions: An alternative to traditional assessment center dimensions? *Personnel Psychology, 47*, 109-121.
- Klimoski, R. J., & Brickner, M. (1987). Why do assessment Centres work? The puzzle of assessment center validity. *Personnel Psychology, 40*, 243-260.
- Kline, P. (1999). *Handbook of Psychological Testing* (2nd ed.). New York: Routledge.
- Kudisch, J. D., Ladd, R. T., & Dobbins, G. H. (1997). New evidence on the construct validity of diagnostic assessment Centres: The findings may not be so troubling after all. *Journal of Social Behavior and Personality, 12*, 129-144.
- Lance, C. E. (2008a). Why assessment Centres do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 84-97.
- Lance, C. E. (2008b). Where have we been, how did we get there, and where shall we go? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 140-146.
- Lance, C. E. (in press). Research into task-based assessment Centres. In D. J. R. Jackson, C. E. Lance & B. J. Hoffman (Eds.), *The psychology of assessment centers*. New York: Routledge.
- Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain't trait it must be method: (Mis)application of the multitrait-multimethod methodology in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences* (pp. 337-360). New York: Routledge.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*, 202-220.
- Lance, C. E., Foster, M. R., Nemeth, Y. M., Gentry, W. A., & Drollinger, S. (2007). Extending the nomological network of assessment center construct validity: Prediction of cross-situationally consistent and specific aspects of assessment center performance. *Human Performance, 20*, 345-362.
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology, 89*, 377-385.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R., & Smith, D. E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance, 13*, 323-353.
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*, 228-244.
- Lievens, F. (1998). Factors which improve the construct validity of assessment Centres: A review. *International Journal of Selection and Assessment, 6*, 141-152.
- Lievens, F., & Christiansen, N. D. (in press). Core debates in assessment center research: Dimensions 'versus' exercises. In D. J. R. Jackson, C. E. Lance & B. J. Hoffman (Eds.), *The psychology of assessment centers*. New York: Routledge.
- Lowry, P. E. (1997). The assessment center process: New directions. *Journal of Social Behavior & Personality, 12*, 53-62.
- Markus, L. H., Cooper-Thomas, H. D., & Allpress, K. N. (2005). Comfounded by competencies? An evaluation of the evolution and use of competency models. *New Zealand Journal of Psychology, 34*, 117-126.
- Mintzberg, H. (1971). Managerial work: Analysis from observation. *Management Science, 18*, 97-110.
- Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper

- & Row.
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, 56, 63-75.
- Pearson, C. A. L., & Chatterjee, S. R. (2003). Managerial work roles in Asia: An empirical study of Mintzberg's role formulation in four Asian countries. *Journal of Management Development*, 22, 694-707.
- Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior & Human Decision Processes*, 38, 76-91.
- Robertson, I. T., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, 55, 171-183.
- Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, 55, 171-183.
- Roth, P. L., Bobko, P., & Mcfarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, 58, 1009-1037.
- Russell, C. J., & Domm, D. R. (1995). Two field tests of an explanation of assessment centre validity. *Journal of Occupational and Organizational Psychology*, 68, 25-47.
- Ruth, D. (2006). Frameworks of managerial competence: Limits, problems and suggestions. *Journal of Managerial Psychology*, 30, 206-226.
- Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401-410.
- Sackett, P. R., & Harris, M. M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology*, 214-229.
- Sackett, P. R., & Lievens, F. (2008). Personnel Selection. *Annual Review of Psychology*, 59, 419-450.
- Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (1999). A new frame for frame-of-reference training: Enhancing the construct validity of assessment Centres. *Journal of Applied Psychology*, 87, 735-746.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Shapira, Z., & Dunbar, R. L. M. (1980). Testing Mintzberg's managerial roles classification using an in-basket simulation. *Journal of Applied Psychology*, 65, 87-95.
- Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Spicer, J. (2005). *Making sense of multivariate data analysis*. London: Sage.
- Stillman, J. A., & Jackson, D. J. R. (2005). A detection theory approach to the evaluation of assessors in assessment centres. *Journal of Occupational and Organizational Psychology*, 78, 581-594.
- Taylor, P. J., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organizations and recruitment firms. *New Zealand Journal of Psychology*, 32, 49-54.
- Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a "hyperdimensional" taxonomy of managerial competence. *Human Performance*, 13, 205-251.
- Thornton, G. C., III., & Krause, D. E. (2009). Selection versus development assessment Centres: an international survey of design, execution, and evaluation. *The International Journal of Human Resource Management*, 20, 478-498.
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372-376.
- Williams, K. M., & Crafts, J. L. (1997). Inductive job analysis: The job/task inventory method. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 51-88). Palo Alto, CA: Davies-Black Publishing.
- Yukl, G. (2010). *Leadership in organizations*. Upper Saddle River, NJ: Pearson Education.
- Korea.

### Corresponding Author:

Duncan J.R. Jackson  
College of Business Administration,  
The University of Seoul  
90 Jeonnong-dong Dongdaemun-  
ku  
Seoul  
South Korea 130-743  
duncan@uos.ac.kr

### Author Note

This paper presents results from a larger study on task-based assessment Centres. The authors would like to thank Brian J. Hoffman, Charles E. Lance, Vikki Andrews, Desley Thompson, and Paul Millin for their contributions to this paper.

This work was supported by the New Professor's Research Fund 2010 from the University of Seoul, South

© This material is copyright to the New Zealand Psychological Society. Publication does not necessarily reflect the views of the Society.