www.AUSTRALIANACADEMICPRESS.com.au

ARTICLE AVAILABLE ONLINE
The Australian and New Zealand Journal of Organisational Psychology

A Critical Examination of the Internal Consistency of Competencies Assessed Across Multiple Methods

Paul Englert[1], Duncan J. R. Jackson[2] and Marco van Gelderen[3]

[1] *Psytech International Limited, Bedfordshire, United Kingdom*

[2] *Department of Management, Massey University Albany, New Zealand*

[3] *Faculty of Economics and Business Administration, Department of Management & Organization, VU Centre for Entrepreneurship, The Netherlands*

Competency models represent a popular concept that is widely applied in such areas as human resource selection, performance management, employee development, and job analysis. Despite widespread popularity and applied use, little is known about the measurement properties of competencies and whether they can reasonably be used to provide meaningful feedback or as a basis for employment decisions. The present study evaluated the internal consistency of multimethod competency scores obtained from a developmental assessment centre. A total of 214 managers participated in this study. Results suggested that there was virtually no support for the idea that multimethod-derived competencies could be regarded as meaningful, internally consistent, underlying characteristics. The article argues that competency models that employ multimethod approaches merely provide a semantic framework for conveying the complexity of a person specification, a framework which is qualitative not quantitative. These findings are discussed in relation to their implications for behavioral assessment in employment.

■ **Keywords:** assessment centre, competency measurement, multi-method assessment, quantitative and qualitative frameworks

Competencies refer to a popular set of variables that are used across a wide array of management settings as performance predictors or criteria (McClelland, 1973; Prahalad & Hamel, 1990; Rothwell & Lindholm, 1999; Schippmann et al., 2000). Although definitions of the term have been criticised as conceptually ambiguous (Grzeda, 2005; McKenna, 2002), Hoffmann's (1999) review presents a useful framework for understanding the diverse applications of competencies. According to Hoffmann (p. 276), three broad definitions exist in research and practice, namely (1) observable performance, (2) the standard of quality of the outcome of the person's performance and (3) the underlying attributes of a person. The first of these definitions is associated with manifest behaviour (Boam & Sparrow, 1992). The second deals with industry standards and meeting quality benchmarks (Grzeda, 2005). The third refers to underlying constructs (e.g., motives, traits, skills, social roles, self-image, or knowledge) that allegedly drive performance (Boyatzis, 1982; Spencer & Spencer, 1993). The scope of the present article will focus on the last, underlying attribute competencies (UACs), definition and issues around the measurement of such constructs.

A previous version of this article was presented at the 23rd annual conference of the Society for Industrial–Organizational Psychology, San Francisco, April 2008. This paper represents part of a larger study into task-based assessment centres.

ADDRESS FOR CORRESPONDENCE: Paul Englert, PO Box 10–914, Wellington, New Zealand. E-mail: paul.englert@opragroup.com

Englert, P., Jackson, D.J.R. & van Gelderen, M. (2011). A critical examination of the internal consistency of competencies assessed across multiple methods. *The Australian and New Zealand Journal of Organisational Psychology*, *4*, 11–19. DOI 10.1375/ajop.4.1.11

Supporting evidence for competency models is primarily qualitative. Most commercial competency models provide a dictionary of terms that create a semantic framework to understand the knowledge, skills and attributes of jobs (for a popular example, see Lombardo & Eichinger, 2002, 2003). Quantitative support for competency models is therefore assumed by practitioners under the assumption that commonly used competency models do indeed have a quantitative structure.

Multiple Methods as the Ultimate Approach to Measuring Competencies

Even within the underlying attribute definition of competencies, attempts to isolate and, in effect, measure such variables are wide and varied. In fact, the very scientific basis for competencies has met with scepticism in the literature (Barrett & Depinet, 1991; Lievens, Sanchez, & de Corte, 2004). There are suggestions that UACs lack empirical evidence (Barrett & Depinet, 1991; Heinsman, de Hoogh, Koopman, & van Muijenaap, 2007) and are overly narrow in their focus (McKenna, 2002, 2004). Part of the confusion in this regard may emanate from the consultant-driven distribution of UACs at the possible cost of scientific rigor (Athey & Orth, 1999). Given their popularity, a better understanding of the measurement properties associated with UACs appears to be a matter of urgency. To date, measurement in this regard has given rise to a number of monomethod and multimethod approaches.

A key example of a monomethod approach to UAC assessment can be seen in Bartram's work on SHL's Great Eight competency framework (Bartram, 2005; Bartram, Robertson, & Callinan, 2002; Kurz & Bartram, 2002). The Great Eight refers to eight broad competencies (e.g., *Leading and Deciding*) that were developed through factor and multiple dimension-scaling analyses (see Kurz & Bartram, 2002), and, thus, carry major advantages vis-à-vis their empirical basis. The instruments associated with this approach are typically used in the criterion domain as multi-faceted indicators of job performance (see Bartram, 2005). However, these instruments are in the form of ratings on questionnaire items and, therefore, represent a monomethod and possibly narrow (or common method) perspective (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Bartram shows evidence for this as a potential problem in his (2005) study where he correlated the Great Eight Competency Framework (numbering > 100 items) with a simple six-item measure of overall job performance (from Nyfield, Gibbons, Baron, & Robertson, 1995). The sample-weighted mean correlation between the competency measures and the simple overall job performance measure was .79, suggesting either (a) redundancy in the Great Eight measures or (b) criterion contamination whereby the predictor competency measures informed on the job performance ratings.

The Society of Industrial and Organizational Psychology (SIOP) taskforce on competency modelling contended that the 'ultimate' approach to competency assessment involves the use of multiple methods (Schippmann et al., 2000). Specifically, they suggest the following approach for a rigorous competency assessment: 'A variable combination and logically selected mix of multiple methods are used to obtain information (e.g., interviews, focus groups, observation, questionnaires), depending on the research setting, target population, and intended application' (p. 716). This description implies a tailored multimethod approach to UAC measurement incorporating a number of different perspectives and methods.

The Link Between Assessment Centers and Quantitative Support for Competencies Measured by Multiple Methods

Given references to UAC measurement and the use of multiple methods, there is an implied link between the quantification of competencies and assessment centres (ACs). As their unit of measurement, traditional approaches to ACs have typically referred to assessment *dimensions* (e.g., Meriac, Hoffman, Woehr, & Fleisher, 2008; Pynes, Bernardin, Benton, & McEvoy, 1988). However, conceptually, the international guidelines on ACs treat UACs and dimensions as equivalent (International Task Force on Assessment Center Guidelines, 2009). Nevertheless, the literature seems to be divided between a research body that deals specifically with competency modelling (see Lievens et al., 2004) and a separate body that deals with ACs (see Bowler & Woehr, 2006).

In terms of points of intersection between the two literatures, Boyatzis (1982), one of the original proponents of the UAC approach, suggests that competency testing is different from, and superior to, AC approaches. In response to this, Barrett and Depinet (1991, p. 1021) state that 'the evidence does not show that competency tests can match the known strengths and validity of assessment centers'. Barrett and Depinet are most likely referring to criterion validity here. However, ACs have a long legacy of measurement problems associated with dimensions (i.e., construct validity internal to ACs, see Bowler & Woehr, 2006; Fleenor, 1996; Lance, 2008a; Sackett & Dreher, 1982). Although there are parallels between the AC findings and the competency literature, the exact extent to which these issues pose a threat to the measurement of UACs is, as yet, unknown.

The SIOP taskforce on competency modeling drew further intersections between ACs and UACs, stating that 'the dimensional structure of assessment centers, and the resulting operational definitions of the broad,

generic individual difference dimensions using behavioral statements, was a portent of things to come in the realm of competency modeling' (Schippmann, et al., 2000, p. 709). Similarly, the applied perspective on UACs includes the use of ACs as potential indicators of competencies (e.g., see Lombardo & Eichinger, 2002, 2003). Questions still remain, however, as to whether there is a difference between 'accepted' competency models (see Tett, Guterman, Bleier, & Murphy, 2000) and sets of AC dimensions. The AC literature has focused almost exclusively on the construct validity of dimensions based on performance in simulation exercises (Bowler & Woehr, 2006; Lance, Lambert, Gewin, Lievens, & Conway, 2004; Lievens & Conway, 2001; Meriac, et al., 2008; Woehr & Arthur, 2003). The SIOP taskforce on competencies, in contrast, recommends the use of quite different methods, and not only those to do with behavioral simulations (Schippmann, et al., 2000, p. 716).

Research Aim

The popularity of competencies implies a desire to substantively assess such variables. However, the competency literature has been criticised for failing to deliver in terms of providing scientific support for such variables (Barrett & Depinet, 1991; Lawler, 1996; Lievens et al., 2004). An obvious gap in this literature body is seen in the lack of research on the use of multiple methods to assess UACs, as recommended by the SIOP taskforce (Schippmann et al., 2000). While the AC literature goes some way towards this end (see Lance, 2008a), the typical focus of this literature body is almost exclusively on patterns observed across simulation exercises (Bowler & Woehr, 2006) and not across other methods (e.g., psychometrics). Two decades ago, Raven (1988) expressed concerns about the psychometric properties associated with competency measures. In particular, he raised questions about the relevance of internally consistent scale scores as meaningful representations of UACs. Raven's concerns bear relevance to the present context now more than ever.

UACs are popular, yet lack empirical research, particularly around the SIOP recommendations for the use of multiple methods. Given that UACs are being used as though they are internally consistent scales for such diverse applications as job analysis, selection, and performance management (Fink, Carr, Phillips, Ruggeberg, & Campion, 2007; Lievens et al., 2004; Lin et al., 2007; Wilson, Levine, Morgeson, Thomsen, & Harvey, 2007), further investigations into their measurement properties are vital. Thus, the general research aim of this study is to present an initial exploration into the internal consistency of a well-established established competency model by examining their application in a multiple-method assessment process.

Method

Participants

A sample of 214 managers participated in a developmental AC, which included supplementary psychometric evaluations. The participating organisation specialised in postal services, insurance, credit, banking, and administrative services. The mean age of participants was 45.53 (SD = 10.33), and gender was split fairly evenly with 54% males and 46% females. Just over half of the participants had completed high school (around 53%) or held a trade certificate or degree (around 25%).

Lominger Competency Framework

The AC in this study was unique in that it used the commercially available Lominger competency framework (Lombardo & Eichinger, 2002, 2003; Tett et al., 2000) as a basis for the UACs assessed in this study. The Lominger library presents a broad array of 119 competencies that can be honed down to a relevant set for each organization (Lombardo & Eichinger, 2002). The means by which these competencies are assessed are largely discretionary in that Lombardo and Eichinger (2002, pp. 211–236) provide suggestions for a broad range of assessment sources and methods. Of these methods, 360-degree approaches and ACs are among those noted by Lombardo and Eichinger, and, in this case, the organisation had determined that AC exercises and psychometrics would be appropriate. The participating organisation had previously indentified 18 Lominger competencies by way of subject matter expert discussion. This same strategy was followed in terms of mapping particular methods (i.e., AC exercises and psychometric evaluations) on to theoretically linked competency areas. This was achieved for each behavioral item in the AC exercises (see below) and each scale from the psychometric assessments. The matching of measures to competencies (see Table 1) involved judgments from teams of HR employees, consultants, and managers.

Assessment Centre Exercises and Assessors

Four exercises were developed, based on the competency model and subject matter expert guidance. The exercises included (a) a group discussion and oral presentation on managing new staff; (b) a group discussion and oral presentation based on selecting new staff; (c) a group analysis exercise that involved the presentation of photos and identifying problems with shop layouts; (d) a role play exercise that involved a coaching discussion with an employee. Each exercise carried an associated 10-item behavioral checklist on which 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

TABLE 1

Multimethod Competency Configuration

#	Lominger Competency Title	Multimethod Configuration	*n(i)*
1	Business acumen	fm; b; ar1; vr1; x1b1; x2b1; x3b1	7
2	Developing direct reports	ue; fl; fa; x4b1	4
3	Informing	ue; fq2; ff; x1b3; x4b2	5
4	Managing and measuring work	fg; ff; x1b6; x4b5	4
5	Listening	fn; fe; x1b7; x2b3; x3b5; x4b6	6
6	Confronting direct reports	fl; md; fn; fo; x4b3	5
7	Drive for results	fq4; spicomp; nr1; x1b4; x2b4; x3b7	6
8	Customer focus	spias; fh; in; ree; x1b5; x2b5; x3b6	7
9	Managing others	me; ue; ce; ff; x2b2; x3b2; x4b4	7
10	Open minded	fq1; x2b6; x3b8	3
11	People smart	fa; me; x3b4; x4b7	4
12	Conflict manager	ue; fe; fn; x4b9	4
13	Cool transactor	md; fc; x4b10	3
14	Agile communicator	vr1; x1b2; x2b10; ex3b10	4
15	Visioning	fq1; x1b10; x2b7; x3b9	4
16	Critical thinking	fl; fa; ue; x4b8	4
17	Solution finder	fm; b; ar1; x1b9; ex2b9; x3b3	6
18	Inspires others	me; ff; fh; x1b8; x2b8	5

Note: *n(i)* = number of components that made up a given competency. 1. Personality-based measures. From the 15FQ+, fm = concrete vs. abstract; b = intellectance; fa = distant aloof vs. empathic; fc = feeling vs. emotionally stable; fe = accommodating vs. dominant; ff = serious vs. enthusiastic; fg = expedient vs. conscientious; fh = retiring vs. socially bold; fl = trusting vs. suspicious; fm = concrete vs. abstract; fn = direct vs. restrained; fo = self-assured vs. apprehensive; fq1 = conventional vs. radical; fq2 = group vs. self-sufficient; fq4 = composed vs. tense driven. From the *Sales Preference Indicator* (spi), spicomp = competitive; spias = adaptive selling; in = integrated networks. From the *Genos Emotional Intelligence* scales, ue = understanding emotions of others; me = managing emotions; ce = controlling emotions; ree = recognising and expressing emotions; md = making decisions. 2. Ability-based methods. From the Psychtec *Graduate Reasoning Test* suite, ar1 = abstract reasoning; vr1 = verbal reasoning; nr1 = numerical reasoning. 3. Performance-based methods. x# = simulation exercises 1 through 4; b# = denotes behavioral checklist items on each simulation exercise.

'Comes up with solutions that have the customer in mind'.

Assessors were ranked a management level above participants (n = 19) or were consultant psychologists (n = 4). The ratio of assessors to participants was 1:2 (see the International Task Force on Assessment Center Guidelines, 2000). Assessor duties were rotated across participants in order to minimize rater error. Frame-of-reference training (FORT) was used (see Lievens, 1998; Pulakos, 1986; Schleicher, Day, Mayes, & Riggio, 1999), involving a two-day period that covered familiarization with rating instruments and exercises, rater errors, and involvement in practice assessments.

Psychometric Evaluations

Two classes of psychometric evaluation were used in this study, comprising personality and ability-based instruments. All of these evaluation tools are commercially available and, as such, detailed information on psychometric properties and subscales associated with each instrument (as listed in Table 1) can be located in the references that follow. Personality-based materials included the 15FQ+ personality inventory (Psytech International, 2002), the sales preference indicator (SPI) (OPRA Consulting Group, 2007), and the Swinburne emotional intelligence test (Palmer & Stough, 2007). Ability-based measures included the verbal, numerical, and abstract reasoning scales from the graduate reasoning test battery (GRT 1) (Psytech International, 1991).

Analyses

The organisation in this study, under the recommendation of a commercially available competency provider, made an attempt to measure UACs as meaningful aggregates to provide developmental feedback to employees. In line with recommendations in the literature (Schippmann et al., 2000), the approach involved forming competency aggregates on the basis of indicators from multiple method. As Raven (1988) implies, if competencies can be considered as meaningful scales, they should return evidence of internal consistency. As such, standardised coefficients alpha (see Kline, 1999; Nunnally & Bernstein, 1994) were estimated for each multimethod competency aggregate along with intercorrelations among UACs.

Results

Table 2 shows standardised coefficients alpha, as estimates of internal consistency, for each Lominger UAC aggregate (Lombardo & Eichinger, 2002, 2003) based on a combination of multiple methods, including personality and ability tests and AC exercises. Alphas were standardized so as to allow for the fact that each of the multimethod competency components was assessed on a different scale (see Kline, 1999). Also shown are the descriptive statistics for each UAC along with their intercorrelations. Despite typically carrying a larger number of items than many AC studies that focus only on exercises (see Bowler & Woehr, 2006), the multimethod UAC measures in this study returned internal consistency estimates that were all below the generally accepted minimum for new measures of .70 (Lance, Butts, & Michels, 2006; Nunnally & Bernstein, 1994). This implies that across the 18 UACs assessed in this study, there was, consistently, little or no agreement among the multimethod item components that were deemed to make up the competency scores. This suggests that there was virtually no support for the idea that competency components clustered together to form meaningful aggregates that could, in turn, be used for practical purposes. The highest internal consistency estimate was associated with the UAC *agile communicator* (alpha = .69) and the lowest was associated with *confronting direct reports* (alpha = .01). The median alpha = .44 (interquartile range = .48), which was well below standards for coefficient alpha (including less stringent standards, see Hair, Anderson, Tatham, & Black, 1998).

Also of note in Table 2 were the intercorrelations among the UACs. Many of these were high and indicative of a degree of measurement redundancy, suggesting, in line with the AC literature, that a lower number of conceptually distinct variables should be assessed (see Gaugler & Thornton, 1989; Lievens, 1998). The mean scores in Table 2 represent UAC aggregates that were used by the organisation for individual difference comparisons and developmental feedback. However, given the low standardised coefficients alpha associated with all UACs, judgments about their measurement, their mean scores, and their intercorrelations are, essentially, meaningless. In summary, there was effectively no evidence that the commercially available Lominger competencies in this study constituted meaningful attributes that could be aggregated across multiple methods.

Discussion

The research aim in this study was to explore whether the multimethod item components of an established and commercially available set of competencies clustered together to form meaningful aggregates (i.e., whether the items that make up competencies cluster meaningfully across methods). Across 18 different Lominger competencies (see Lombardo & Eichinger, 2002; Tett, et al., 2000), strikingly consistent findings emerged in that we essentially found no support for the idea that any of the multimethod UACs in this study could be considered coherent as determined by meaningful, internally consistent scales (see Table 2). With reference to competencies in their *multimethod* form, it is possible that such variables should not be considered as quantitative scales or underlying psychological constructs but as nothing more than the qualitative illustrations (Raven, 1988). If the intention is to assess multimethod UACs as though they are meaningful quantitative aggregates, then for reasons around ethics, fairness, and sensibility, the results of this study suggest that UAC scores should be analysed a priori to show evidence for their basic, underlying psychometric properties (e.g., an assessment of structural evidence, reliability, and validity, see Brannick, 2008).

While, in contrast, *monomethod* competency approaches (e.g., questionnaire-based competency measures) are likely to provide internally consistent scales, their downside is that they may provide a narrow focus and may be redundant with simple overall performance measures (see Bartram, 2005). The multimethod approach, on the other hand, provides a broad focus; however, the results of this study suggest that treating multimethod-derived UACs as aggregates is likely to be ambitious at best. Thus, if the multimethod approach is the preferred model (Schippmann, et al., 2000), then multimethod competency aggregations should, in the very least, be assessed for basic psychometric evidence or they should not be used at all. This could include, for example, evidence for a meaningful factor structure.

It is perhaps the case that competency or AC dimension titles (e.g., *business acumen*, *drive for results*, or *customer focus*) encourage users of assessment procedures to think of such titles as describing scale aggregates (Jackson, Barney, Stillman, & Kirkley, 2007; Jackson, Stillman, & Atkins, 2005). The results of the present study are reminiscent of Sacket and Dreher's (1982, p. 409) AC dimension findings where they stated that they found 'virtually no support for the view that the assessment center technique generates dimensional scores that can be interpreted as representing complex constructs such as leadership, decision making, or organizational acumen'. These results remain as a source of contemporary debate in the AC literature (Lance, 2008a) and the results of the present study suggest a crossover into the competency literature. Also, over and above the AC findings, our results suggest that the failure of multimethod UACs to emerge as internally consistent scales generalises beyond the use of simulation exercises and into the (truly) multimethod domain (see Lance, Baranik, Lau, & Scharlau, 2008).

TABLE 2

Standardized Coefficients Alpha and Correlations among Competency Aggregates with Descriptive Statistics

#	Competency	M	SD	n(i)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	BA	4.06	1.14	7	**.63**																	
2	DDR	4.22	1.51	4	.42	**.43**																
3	INF	4.61	1.12	5	.48	.64	**.09**															
4	MMW	4.25	1.16	4	.44	.53	.74	**.24**														
5	LIS	5.25	0.89	6	.36	.53	.45	.31	**.34**													
6	CDR	4.68	1.48	5	.28	.66	.53	.48	.28	**.01**												
7	DFR	4.42	1.22	6	.68	.44	.47	.40	.36	.38	**.47**											
8	CF	4.32	0.99	7	.70	.44	.47	.44	.38	.30	.57	**.53**										
9	MO	4.24	1.20	7	.60	.64	.57	.48	.56	.41	.65	.67	**.67**									
10	OM	4.05	1.20	3	.63	.36	.28	.19	.32	.25	.54	.55	.53	**.23**								
11	PS	4.40	1.21	4	.47	.52	.47	.35	.49	.31	.46	.58	.64	.37	**.45**							
12	CM	4.37	1.45	4	.36	.69	.53	.44	.57	.56	.39	.45	.63	.33	.57	**.12**						
13	CT	4.96	1.39	3	.34	.56	.40	.29	.40	.54	.32	.31	.46	.22	.45	.61	**.06**					
14	ACOM	4.48	1.31	4	.78	.46	.52	.47	.44	.31	.69	.63	.68	.51	.51	.42	.41	**.69**				
15	VIS	4.01	1.23	4	.76	.43	.43	.37	.38	.33	.67	.68	.63	.80	.45	.40	.32	.64	**.48**			
16	CTK	4.01	1.41	4	.39	.70	.53	.46	.41	.51	.40	.42	.60	.30	.55	.65	.57	.45	.39	**.11**		
17	SF	4.07	1.11	6	.83	.43	.49	.44	.40	.31	.73	.70	.70	.58	.51	.35	.32	.76	.74	.41	**.58**	
18	IO	4.48	1.40	5	.69	.51	.52	.45	.43	.38	.70	.68	.79	.55	.54	.49	.41	.68	.69	.48	.73	**.65**

Note: BA = business acumen; DDR = developing direct reports; INF = informing; MMW = managing and measuring work; LIS = listening; CDR = confronting direct reports; DFR = drive for results; CF = customer focus; MO = managing others; OM =open minded; PS = people smart; CM = conflict manager; CT = cool transactor; ACOM = agile communicator; VIS = visioning; CTK = critical thinking; SF = solution finder; IO = inspires others. Correlations between .24 and .25 significant at $p < .05$. Correlations > .25 significant at $p < .01$. Standardized coefficients alpha appear in bold in the diagonal (median alpha = .44; interquartile range = .48). $N = 214$; $n(i)$ = number of components, treated as items, making up a given composite.

We argue that the competency model in this study, while accepted commercially (Lombardo & Eichinger, 2002) and in the research literature (Tett, et al., 2000), is actually nothing more than a hypothetical and qualitative framework. It is not, however, a set of characteristics that can be considered in the same manner as psychometric scales. The multimethod subcomponents of the competencies in Table 1 included personality scales, ability scales, and behavioural checklists from AC exercises. In and of themselves, such measures may take years to build in terms of the scale development work that is needed to verify reliability and validity. Recently, the AC approach has been criticised as failing to acknowledge the importance of basic psychometrics (Brannick, 2008). It appears that multimethod competency models may also need to take a leaf out of this book. If developmental feedback is being provided to employees on the basis of UACs, then such variables are being assumed to operate as internally consistent and meaningful scales. None of the competencies in this study were internally consistent, providing virtually no support for the idea that they could be used as scales on which to base developmental feedback or guidance for employment purposes.

Our contention is that a separation needs to be drawn between substantive measurement and the formation of competency titles. The mere presence of a competency title does not imply that a competency exists as an internally consistent scale or even as an underlying psychological construct. As Raven (1988) suggests, a better use of such titles might be to provide theoretical or qualitative guidance. For example, competency titles could help to guide the choice of actual assessment tools. However, we are concerned that the temptation for practitioners might be to mount attempts to measure UACs directly and that this desire may supersede basic sensibilities around measurement (see Brannick, 2008; Gorham, 1978; Wernimont & Campbell, 1968).

Our aims in this study were centered on a basic approach to psychometrics and on issues that precede the use of more complex techniques that can be applied to psychometric data (e.g, structural equation modeling, generalisability theory). Other studies have investigated the application of such techniques to analogous datasets (Hoffman, Lance, Bynum, & Gentry, in press; Hoffman & Woehr, in press; Jackson, et al., 2005). Our concerns here, however, relate to those raised by Bannick (2008), in that many practitioners are not even seeking very basic evidence for the structure of their measures, let alone using more complex techniques. In this regard, it has been argued that rudimentary psychometric evidence should be established, particularly before the application of covariance modeling (Brown, 2006) and, more generally, prior to the application of most multivariate approaches to analysis (Tabachnick & Fidell, 2001). In addition, we recognise that a different configuration of measures and a different set of exercises could have resulted in conclusions that differ from those presented here. However, our overall conclusions mirror many of those in the literature that have focused on different assessment configurations and have used more advanced techniques to analyse data (Lance, 2008a, 2008b; Lance, et al., 2004).

Competencies have a pragmatic role in that they cluster the complexity of skills, knowledge, and attributes required or desired for a job into simple to understand dimensions. This has the effect of creating a simple framework of human behaviors. This framework becomes a semantic code for a business to outline the general expectations required of people. However, as indicated in this study, competencies are likely to be *qualitative* and do not appear to have the internally consistent, quantitative structure that is so often assumed by those who use competencies to inform business decisions. Thus, the application of competencies in a business is likely to constitute the transmission of information and standards at a broad level not the alleged fine-grained measurement of behavior such as in assessment. Where competencies are applied qualitatively there is likely to be few problems. Where they are assumed to have quantitative psychometric properties, this study indicates that problems may occur.

References

Athey, T.R., & Orth, M.S. (1999). Emerging competency methods for the future. *Human Resource Management, 38*, 215–226.

Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist, 46*, 1012–1024.

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*, 1185–1203.

Bartram, D., Robertson, I. T., & Callinan, M. (2002). Introduction: A framework for examining organizational effectiveness. In I.T. Robertson, M. Callinan & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 1–10). Chichester, UK: Wiley.

Boam, R., & Sparrow, P. (1992). *Designing and achieving competency*. London: McGraw-Hill.

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.

Boyatzis, R.E. (1982). *The competent manager: A model for effective performance*. New York: Wiley.

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.

Fink, A.A., Carr, L.S., Phillips, G.M., Ruggeberg, B.J., & Campion, M.A. (2007). *Top-rated practice forum: Doing com-*

petencies well. Paper presented at the 22nd Annual SIOP Conference, New York.

Fleenor, J.W. (1996). Constructs and developmental assessment centres: Further troubling empirical findings. *Journal of Business and Psychology, 3*, 319–335.

Gaugler, B.B., & Thornton, G.C. (1989). Number of assessment dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74*, 611–618.

Gorham, W.A. (1978). Federal executive agency guidelines and their impact on the assessment center method. *Journal of Assessment Center Technology, 1*, 2–8.

Grzeda, M.M. (2005). In competence we trust? Addressing conceptual ambiguity. *Journal of Managerial Psychology, 24*, 530–545.

Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate data analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Heinsman, H., de Hoogh, A.H.B., Koopman, P.L., & van Muijenaap, J.J. (2007). Competencies through the eyes of psychologists: A closer look at assessing competencies. *International Journal of Selection and Assessment, 15*, 412–427.

Hoffman, B.J., Lance, C., Bynum, B., & Gentry, B. (in press). Rater source effects are alive and well after all. *Personnel Psychology*.

Hoffman, B.J., & Woehr, D.J. (in press). Disentangling the meaning of multisource feedback: An examination of the nomological network surrounding source and dimension factors. *Personnel Psychology*.

Hoffmann, T. (1999). The meanings of competency. *Journal of European Industrial Training, 23*, 275–285.

International Task Force on Assessment Center Guidelines. (2000). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management, 29*, 315–331.

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., 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 centers: A psychometric comparison. *Human Performance, 18*, 213–241.

Kline, P. (1999). *Handbook of Psychological Testing* (2nd ed.). New York: Routledge.

Kurz, R., & Bartram, D. (2002). Competency and individual performance: Modeling the world of work. In I.T. Robertson, M. Callinan & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 227–255). Chichester, UK: Wiley.

Lance, C.E. (2008a). Why assessment centers 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., Baranik, L.E., Lau, A.R., & Scharlau, E.A. (2008). 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: Received doctrine, verity, and fable in organizational and social research*. 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., 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.

Lawler, E. (1996). Competencies: A poor foundation for the new pay. *Compensation and benefits review, 28*, 20–22.

Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment, 6*, 141–152.

Lievens, F., & Conway, J.M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*, 1202–1222.

Lievens, F., Sanchez, J.I., & de Corte, W. (2004). Easing the inferential leap in competency modeling: The effects of task-related information and subject matter expertise. *Personnel Psychology, 57*, 881–904.

Lin, L., Bolton, K.R., Dunleavy, E.M., Foster, C., Gonder, M.L., Malley, L.A. (2007). *Challenges and future direction of job analysis approaches*. Paper presented at the 22nd Annual SIOP Conference, New York.

Lombardo, M.M., & Eichinger, R. W. (2002). *The leadership machine* (3rd ed.). Mineapolis: Lominger Limited Inc.

Lombardo, M.M., & Eichinger, R.W. (2003). *FYI: For your improvement* (3rd ed.). Mineapolis: Lominger Limited Inc.

McClelland, D.C. (1973). Testing for competence rather than for 'intelligence'. *American Psychologist, 28*, 1–14.

McKenna, S. (2002). Can knowledge of the characteristics of 'high performers' be generalised? *Journal of Managerial Psychology, 21*, 680–701.

McKenna, S. (2004). Predispositions and context in the development of managerial skills. *Journal of Managerial Psychology, 23*, 664–677.

Meriac, J.P., Hoffman, B.J., Woehr, D.J., & Fleisher, M.S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*, 1042–1052.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Nyfield, G., Gibbons, P.J., Baron, H., & Robertson, I.T. (1995, May). *The cross cultural validity of management assessment methods.* Paper presented at the 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando.

OPRA Consulting Group. (2007). *Sales preference indicator technical manual v. 5.1*. Wellington, New Zealand: OPRA Publishing and Consulting Group Ltd.

Palmer, B., & Stough, C. (2007). *Swinburne University emotional intelligence test technical manual v. 2*. Melbourne, Australia: Swinburne University of Technology.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Prahalad, C.K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review, 68*, 79–91.

Psytech International. (1991). *General, critical & graduate test battery: The technical manual*. Pulloxhill, Bedforshire: Published Psychometrics Ltd.

Psytech International. (2002). *15FQ+ fifteen factor questionnaire technical manual.* Pulloxhill, Bedforshire: Published Psychometrics Ltd.

Pulakos, E.D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior & Human Decision Processes, 38*, 76–91.

Pynes, J., Bernardin, H.J., Benton, A.L., & McEvoy, G.M. (1988). Should assessment center dimension ratings be mechanically-derived? *Journal of Business and Psychology, 2*(3), 217–227.

Raven, J. (1988). Toward measures of high-level competencies: A re-examination of McClelland's distinction between needs and values. *Human Relations, 41*, 281–294.

Rothwell, W.J., & Lindholm, J.E. (1999). Competency identification, modelling and assessment in the USA. *International Journal of Training and Development, 3*, 90–105.

Sackett, P.R., & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67*, 401–410.

Schippmann, J.S., Ash, R.A., Battista, M., Carr, L., Eyde, L.D., Hesketh, B. (2000). The practice of competency modeling. *Personnel Psychology, 53*, 703–740.

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 centers. *Journal of Applied Psychology, 87*, 735–746.

Spencer, L.M., & Spencer, S.M. (1993). *Competence at work: Models for superior performance*. New York: Wiley.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Pearson.

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.

Wernimont, P.F., & Campbell, J.P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372–376.

Wilson, M.A., Levine, J.D., Morgeson, F.P., Thomsen, D.J., & Harvey, R.J. (2007). *Unanswered Questions in Competency Modeling and Job Analysis*. Paper presented at the 22nd Annual SIOP Conference, New York.

Woehr, D.J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231–258.