Content Validation: The Forgotten Step-child or a Crucial Step in AC Validation?

Klaus Müller & Gert Roodt

Agenda

- Background
- Our Perspective
- Research Objectives
- Methodology
- Results
- Conclusions
- Proposed Future Research
- Practical Impact

UNIVERSITY OF JOHANNESBURG

Purpose of Presentation

- To demonstrate the utility of content analysis in the assessment of AC validity
- To recommend future research on this topic
- To show the practical application and implications of the methods used to establish AC content validity

UNIVERSITY OF JOHANNESBURG

Background

- Approx. ⅓ of content in ACs developed abroad & imported for local use (Krause, Rossberger, Dowdeswell, Venter, & Joubert, 2011)
- AC content often borrowed from the USA & UK
- For evaluating AC effectiveness, content validation is used in 28% of cases in Africa (Hughes et al, 2012)
- Very little published research found on AC content validation
- Preference given to construct validation
- Content validity often assumed or erroneously implied

UNIVERSITY OF JOHANNESBURG

Background continued

- AC participant presented with contextual stimulation that requires action from participant
- "…active ingredient…" in ACs the use of exercise-simulations and associated dimensions (Hoffman et al., 2011, p. 380)
- Content exercise validation very popular in the medical training fraternity
- Importance of AC simulation content-culture fit

UNIVERSITY OF JOHANNESBURG

Our Perspective

- Content validation NB building-block for achieving construct and criterion validity
- Statistical procedures totally dependent on **quality** of data entered
- Case of “garbage in equals garbage out”
- Performing content analysis before construct validity will increase overall validity
- Content validation suited to analysing relevance of AC simul in a cultural context
- Experts are qualified to judge AC content for domain relevance & applicability

UNIVERSITY OF JOHANNESBURG

Our Perspective Continued

Def. *Content validity is the quantitative and qualitative evaluation by experts of the relevance and representativeness of an assessment measure with regard to the targeted sample domain* (Muller & Roodt, 2012)

Experts evaluate and judge AC content on:

- Relevance of a competency wrt specified job;
- Representativeness of the VAC simulations wrt specified job;
- Comprehensiveness with which the VAC assesses competencies;
- Elements that could disadvantage a particular demographic group

UNIVERSITY OF JOHANNESBURG

Research Objectives

- To investigate whether the Virtual Assessment Centre (VAC) is content valid
- To examine the degree of agreement between subject matter experts (SMEs) and experienced managers (FEs) on simulation content
- To establish whether the VAC content contains elements that adversely impact any demographic group

UNIVERSITY OF JOHANNESBURG

Methodology

- A purposeful sampling approach used in selecting expert participants (N=22)
- NB that experts were those whose experience and knowledge relate to the topic (ACs and managerial-related competencies)
- Industrial Psychologists & Managers

UNIVERSITY OF JOHANNESBURG

Methodology – Participants

Table 1: *Characteristics of Participants*

Item	Category	SMEs		FEs		Combined	
		Freq	%	Freq	%	Freq	%
Age	25-34	1	7	4	44,4	5	22,7
	35-44	4	30,7	5	55,6	9	40,9
	45-54	4	30,7	0	0	4	18,2
	55-64	4	30,7	0	0	4	18,2
Gender	Male	5	38	6	66,6	11	50
	Female	8	52	3	33,3	11	50
Language	English	4	30,7	8	88,8	12	54,5
	Afrikaans	8	52	1	11,2	9	40,9
	IsiXhosa	1	0,7	0	0	1	4,5
Total		13		9		22	100

UNIVERSITY OF JOHANNESBURG

Methodology – Instrument

Seven dimensions were identified from literature :

- ***Competency Area – Job Correspondence*****.** 5 sub-dimensions relating to job of manager: CT, PL, COMM, TPM, & CFOC
- ***Job Competency-Simulation Match*****.** Extent to which VAC simulations match the job competencies.
- ***Complexity*****.** Level of the VAC content matches the selected job complexity
- ***Fidelity*****.** Assess and compare realism of VAC simulations to real-life, work related scenarios

UNIVERSITY OF JOHANNESBURG

Methodology – Instrument cont.

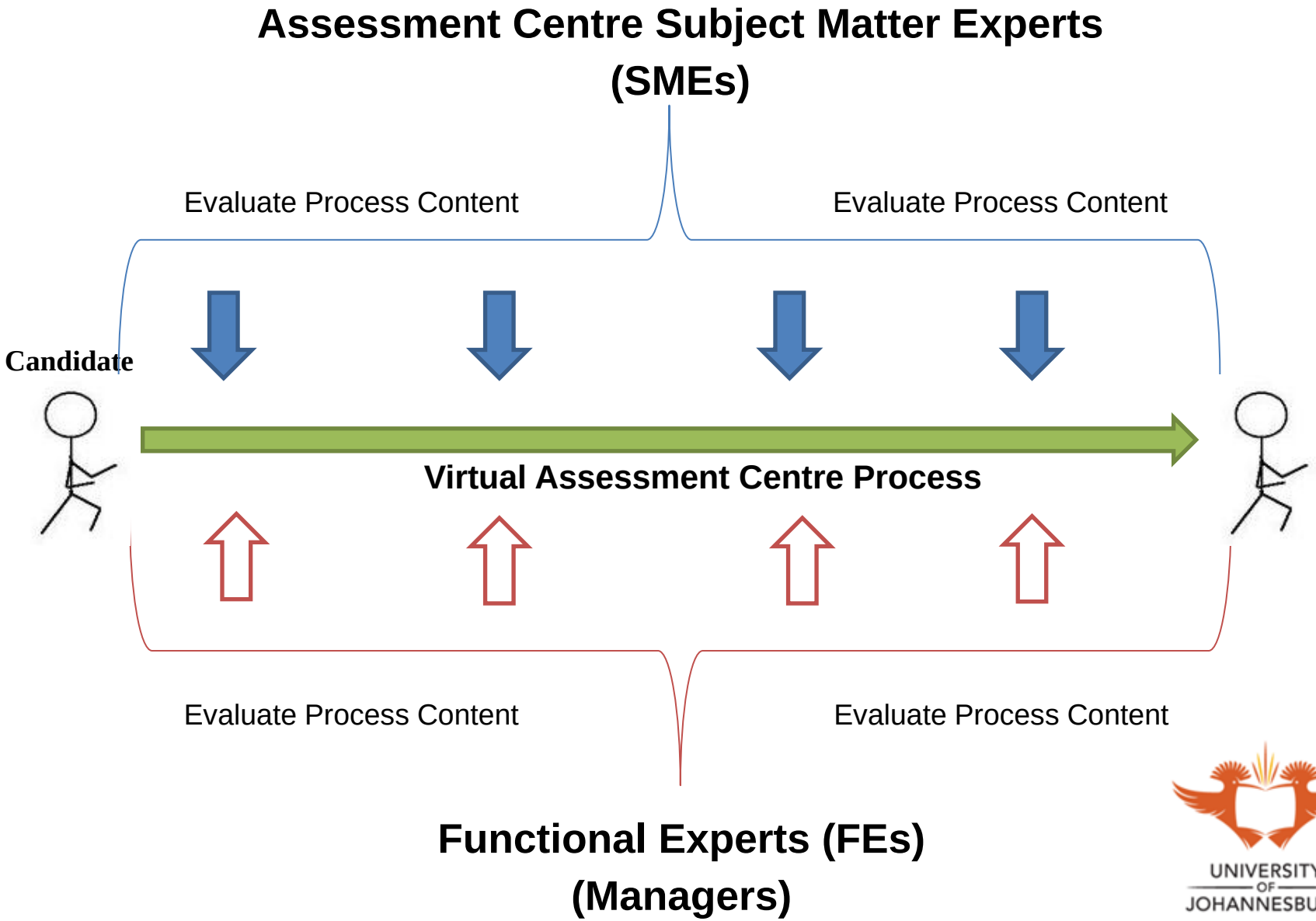
- ***Adverse Impact*.** In SA, NB to assess & address concerns of adverse impact & that all demographics have potentially similar selection ratios.
- ***Economic Considerations*.** Potential value & effectiveness of VAC assessed in order to determine relative ECO worth of VAC
- ***Ethical Considerations*.** Possible & potential ethical issues of VAC were addressed.

UNIVERSITY OF JOHANNESBURG

Methodology – Research Procedure

- Research procedure that was implemented was informed by best practices (*cf.* Haynes et al., 1995; Kolk, Born & van der Flier, 2002; Sackett, 1987; Sproule, 2009; Yagmaie, 2003)
- Guided by procedures from content validation (*cf.* Polit, Beck, & Owen, 2007; Rubio, Berg-Weger, & Tebb, 2003) & cross-cultural adaptation studies (*cf.* Brandt, 2005; Mosdell, Dunnette, & Ameen, 2010; van de Vijver & Tanzer, 2004).
- Issues regarding adverse impact were informed and guided by practices & procedures adapted from previous studies (*cf.* Bosco & Allen, 2012; De Corte, Sackett, & Lievens, 2010; Risavy & Hausdorf, 2011)

Methodology – Process



UNIVERSITY OF JOHANNESBURG

Methodology – Process cont

- Both types of experts view & evaluate same assessment process
- Virtual aspect → method of content delivery via IT interface
- Audio via telephone (linked via conference call)
- Experts visually see user input & hear dialog between candidate & assessors
- Experts tasked with completing 50 item evaluation schedule

UNIVERSITY OF JOHANNESBURG

Methodology – Statistical Analysis

- ICC consistency model (Shrout & Fleiss, 1979) to determine levels of consistency (reliability) of raters
- Communalities for raters on a single factor calculated to ascertain how much raters had in common
- To ascertain the loadings of each rater on the common factor
- Equivalent to the correlations of each raters' ratings with the common factor (Uebersax, 2000)
- Mann-Whitney U test to determine group differences
- Z-scores to establish if group scores of SMEs & FEs on dimensions were within 1 SD

UNIVERSITY OF JOHANNESBURG

Results - Descriptives

Table 2: Means, Confidence Intervals, Variances, Standard Deviations

Dimension	*X*	95% CI		S^2	*SD*	*Min*	*Max*
1. CAJC		LL	UL				
1.1 CT	5.69	5.32	6.07	0.72	0.85	3.67	7
1.2 PL	4.98	4.5	5.46	1.17	1.08	2.33	7
1.3 COMM	5.42	4.93	5.92	1.25	1.12	3	7
1.4 TPM	5.41	5	5.81	0.84	0.91	3.67	7
1.5 CFOC	5.44	4.89	5.99	1.53	1.24	3	7
2. JCM	5.53	5.16	5.9	0.7	0.84	4	6.8
3. CMPLX	5.4	5.06	5.74	0.59	0.77	3.6	6.6
4. FDL	5.59	5.59	6.07	1.16	1.07	3.14	7
5. AI	6.49	6.23	6.71	0.35	0.59	5.14	7
6. ECO	4.8	4.47	5.14	0.57	0.75	3.29	6.43
7. ETH	6.34	5.99	6.69	0.63	0.79	4.75	7

UNIVERSITY OF JOHANNESBURG

Results - Descriptives

- Overall, mean scores higher than scale mid-point 4
- Suggests slightly neg skewed score distribution Mean scores relatively high, relatively sml SDs
- Particularly high scores for AI (6.49) & ETH (6.34)
- Indicates that most raters agree that dimensions relating to VAC CV is high
- This is the first portion of proving VAC content validity, next we assess rater consistency

UNIVERSITY OF JOHANNESBURG

Results – Rater Consistency

Table 3: *ICC (Consistency Model) Expert Dimension Comparison*

Dimension	Expert	ICC		F	df	p
		Single α	Avg α			
1.1 CT	SME	.54	.78	4.49	12	.01
	FE	.59	.81	5.33	8	.01
	Overall	.61	.82	5.63	21	.01
1.2 PL	SME	.68	.87	7.503	12	.01
	FE	.86	.95	19.19	8	.01
	Overall	.78	.92	11.88	21	.01
1.3 COMM	SME	.66	.86	6.92	12	.01
	FE	.85	.94	17.70	8	.01
	Overall	.79	.90	9.88	21	.01
1.4 TPM	SME	.66	.87	6.93	12	.01
	FE	.80	.92	13.00	8	.01
	Overall	.77	.91	11.16	21	.01
1.5 CFOC	SME	.80	.92	13.14	12	.01
	FE	.89	.96	25.71	8	.01
	Overall	.89	.96	25.31	21	.01
2. JCM	SME	.50	.83	6.02	12	.01
	FE	.45	.81	5.16	8	.01
	Overall	.55	.86	7.02	21	.01
3. CMPLX	SME	.43	.79	4.70	12	.01
	FE	.30	.69	3.17	8	.01
	Overall	.36	.74	3.77	21	.01

Results – Rater Consistency cont

Dimension	Expert	*ICC*		*F*	*df*	*p*
		Single a	*Avg a*			
4. FDL	**SME**	**.46**	**.86**	**6.91**	**12**	**.01**
	FE	**.60**	**.91**	**11.32**	**8**	**.01**
	Overall	**.63**	**.92**	**12.95**	**21**	**.01**
5. AI	**SME**	**.20**	**.64**	**5.41**	**12**	**.01**
	FE	**.60**	**.98**	**11.60**	**8**	**.01**
	Overall	**.38**	**.82**	**5.41**	**21**	**.01**
6. ECO	**SME**	**.21**	**.64**	**2.81**	**12**	**.01**
	FE	**.16**	**.58**	**2.35**	**8**	**.03**
	Overall	**.21**	**.65**	**2.88**	**21**	**.01**
7. ETH	**SME**	**-.03**	**-.15**	**0.87**	**12**	**.58**
	FE	**.49**	**.77**	**4.38**	**8**	**.01**
	Overall	**.15**	**.41**	**1.69**	**21**	**.57**
OVERALL		**.35**	**.96**	**24.95**	**21**	**.01**

Note. ETH SME value negative due to negative average covariance amongst items.

UNIVERSITY OF JOHANNESBURG

Results – Rater Consistency cont

- ICC consist coef. for SMEs & FEs largely congruent
- Visible diff in consistency betw. SMEs & FEs for CMPLX (*ICC* = .79, *ICC* = .69), & AI (*ICC* = .64, *ICC* = .98)
- Low neg ICC value for SMEs on ETH (*ICC* = -.15) due to neg avg covariance amongst items
- Single ICC is an index for consistency (reliability) of ratings for one, typical, single rater
- FEs typically scored, on avg, higher on dimension consistency than SMEs

UNIVERSITY OF JOHANNESBURG

Results – Factor Analysis? – YES!

- A single common factor can be interpreted as degree that raters' ratings are associated with the latent factor (VAC process content)
- Latent common factor is not truly the same as a construct that is being measured.
- Rather, a factor is inferred from raters' imperfect perceptions or a shared interpretation of VAC content validity. (Uebersax, 1993)

UNIVERSITY OF JOHANNESBURG

Results – Factor Analysis

Table 4: Rater Communalities and Single Rater vs Group Correlation

Rater	Communalities	Rater vs. Group Correlation
R1	.84	.45
R2	.73	.41
R3	.60	.27
R4	.90	.46
R5	.55	.32
R6	.71	.19
R7	.86	.41
R8	.90	.49
R9	.65	.30
R10	.76	.42
R11	.80	.22

UNIVERSITY OF JOHANNESBURG

Results – Factor Analysis cont

Rater	Communalities	Rater vs. Group Correlation
R12	.74	.33
R13	.80	.43
R14	.75	.29
R15	.69	.21
R16	.89	.48
R17	.74	.32
R18	.94	.47
R19	.58	.35
R20	.71	.25
R21	.76	.35
R22	.86	.26

Extraction Method: Maximum Likelihood. Chi-Square (400.97), DF (209), Sig $p \leq .01$

UNIVERSITY OF JOHANNESBURG

Results – Factor Analysis cont

- Majority of raters' communalities were high, range of .55 to .94.
- Most rater communalities .70 >
- Comparatively, low communalities for Rater 5 (.55) & Rater 19 (.58)
- Single-rater correlation with overall rater group score ranged from .19 to .49
- Comparatively, low to moderate correlations were found for R3 (.19), R11 (.22), R15 (.21), and R20 (.25).
- 68% of rater group correlations > = .30

UNIVERSITY OF JOHANNESBURG

Results – Group Comparison

Table 5: Mann-Whitney U Test of SMEs (n=13) and FEs (n=9) diff on dimensions

Dimension	*Mann-Whitney U*	*Wilcoxon W*	*z*	*p*
1.CAJC				
1.1 CT	**34.00**	**79.00**	**-1.67**	**.09**
1.2 PL	**30.50**	**75.50**	**-1.89**	**.06**
1.3 COMM	**43.50**	**88.50**	**-1.01**	**.31**
1.4 TPM	**24.00**	**69.00**	**-2.33**	**.02***
1.5 CFOC	**24.50**	**69.50**	**-2.29**	**.02***
2. JCM	**23.50**	**68.50**	**-2.39**	**.02***
3. CMPLX	**44.00**	**89.00**	**-0.98**	**.33**
4. FDL	**22.50**	**67.50**	**-2.41**	**.02***
5. AI	**57.50**	**148.50**	**-0.07**	**.95**
6. ECO	**26.00**	**71.00**	**-2.18**	**.03***
7. ETH	**42.50**	**133.50**	**-1.11**	**.27**

Note. *$p \leq .05$ (2-tailed)

UNIVERSITY OF JOHANNESBURG

Results – Group Comparison cont

- Exists some statistically significant differences between SMEs n=13 & FEs n=9
- TPM (U = 24; p = .02); CFOC (U = 24,5; p = .02); JCM (U = 23,5; p = .02); FDL (U = 22,5; p = .02); ECO (U = 26; p = .03)
- 5 dimensions exhibited mostly small but significant differences

UNIVERSITY OF JOHANNESBURG

Results – Group Comparison cont

Table 6: Z-Scores of SMEs and FEs on Dimensions on Evaluation Schedule.

Dimension	$\bar{x}$ SME	$\bar{x}$ FE	($\bar{x}$SME+ $\bar{x}$FE) / 2	SD(SME)	SD(FE)	Z-SME	Z-FE	|Z-SME| +|Z-FE|
1. CAJC								
1.1 CT	5.97	5.3	5.63	0.66	0.96	0.51	-0.34	0.85
1.2 PL	5.33	4.48	4.90	0.89	1.17	0.47	-0.36	0.83
1.3 COM	5.64	5.11	5.37	0.98	1.28	0.27	-0.20	0.47
1.4 TPM	5.77	4.89	5.33	0.79	0.85	0.55	-0.51	1.06
1.5 CFOC	5.97	4.67	5.32	0.89	1.29	0.73	-0.50	1.23
2. JCM	5.88	5.02	5.45	0.63	0.86	0.68	-0.5	1.18
3. CMPLX	5.57	5.16	5.36	0.68	0.85	0.30	-0.24	0.54
4. FDL	6.08	4.89	5.48	0.72	1.15	0.82	-0.51	1.33
5. AI	6.55	6.4	6.47	0.46	0.76	0.16	-0.09	0.25
6. ECO	5.05	4.44	4.74	0.67	0.74	0.45	-0.41	0.86
7. ETH	6.54	6.06	6.3	0.62	0.95	0.38	-0.25	0.63

Results – Group Comparison cont

- SMEs & FEs differ on z-scores on all dimensions
- |Abs| values on dims CT (0.85), PL (0.83), COM (0.47), CMPLX (0.54), AI (0.25), ECO (0.86), & ETH (0.63) within range of 1 SD
- Fair degree of similarity amongst SME and FE scores.
- However, for dims TPM (1.06), CFOC (1.23), JCM (1.18), and FDL (1.33), z-scores diff outside range of 1 SD
- Can be interpreted as moderate to slight differences on dimensions wrt SMEs and FEs

UNIVERSITY OF JOHANNESBURG

Conclusions

- To infer a level of VAC content validity, it is 1stly required that raters rate relevant dimensions highly
- 2ndly, required that raters show a high level of consistency in dimension rating
- A mean score of 5 (out of 7) and larger is taken to indicate a high score.
- Overall, the majority of mean scores for dimensions are high, with relatively small SD
- Indicate that most raters agree that VAC content valid

UNIVERSITY OF JOHANNESBURG

Conclusions cont

- FA results indicate a high degree of rater communality with a single factor
- Rater consistency -> high degree of inter-rater agreement amongst experts
- Visual inspection of ETH showed extremely high item scores (little possible variance)
- Consistently slightly lower reliabilities found for SMEs' ratings compared with FEs
- Indicating that FEs slightly more consistent in scoring In general

UNIVERSITY OF JOHANNESBURG

Proposed Future Research

- Use procedures and approaches to prove content validity of an AC
- Sound the return of content analysis as a useful method for establishing content validity in a scientific manner
- Increase the quality of substantive content before construct validity is conducted
- NB of face validity

UNIVERSITY OF JOHANNESBURG

Practical Impact

- Increase quality of assessment measures
- Content created abroad can be assessed i.t.o validity (relevance & applicability) for use in South Africa
- Solid method to evaluate content-culture fit
- Content validation can positively increase perceptions of assessment fairness
- Increase face validity -> more likely to find acceptance amongst managers & participants

UNIVERSITY OF JOHANNESBURG

Questions?

UNIVERSITY OF JOHANNESBURG

Contact Details

Gert Roodt groodt@uj.ac.za

Klaus Müller klaus@m-network.co.za

UNIVERSITY OF JOHANNESBURG

Reference List

Bosco, F. A., & Allen, D. G. (2012). Executive attention as a predictor of employee performance: Reconsidering the relationship between cognitive ability and adverse impact potential. Manuscript under review at *Journal of Applied Psychology.* doi: 10.5464.AMBPP.2011.191.a

Brandt, A. (2005). Translation, cross-cultural adaptation, and content validation of the QUEST. *Technology and Disability, 17,* 205-216.

De Corte, W., Sackett, P., & Lievens, F. (2010). Selecting predictor subsets: Considering validity and adverse impact. *International Journal of Selection and Assessment, 18*(3), 260-270.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247.

Hughes, D., Riley, P., Shalfrooshan, A., Gibbons, A., & Thornton, G. (2012). A Global Survey of Assessment Centre Practices. The a&dc Group.

Krause, D. E., Rossberger, R. J., Dowdeswell, K., Venter, V., & Joubert., T. (2011). Assessment Center Practices in South Africa. *International Journal of Selection and Assessment, 19*(3), 262-275

Kolk, N., Born, M., van der Flier, H., Olman, J. M. (2002). Assessment center procedures: Cognitive load during the observation phase. *International Journal of Selection and Assessment, 10*(4), 271-278. doi: 10.1111/1468-2389.00217

Mosdell, J, Balchin, R., & Ameen, O. (2010). Adaptation of aphasia tests for neurocognitive screening in South Africa. *South African Journal of Psychology, 40*(3), 250-261.

Muller, K. P., & Roodt, G. (2012). *The Content Validation of a Virtual Assessment Centre.* Unpublished master's dissertation, University of Johannesburg, Gauteng, South Africa.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30,* 459-467. doi: 10.1002/nur.20199

Risavy, S., & Hausdorf, P. A. (2011). Personality testing in personnel selection: Adverse impact and differential hiring rates. *International Journal of Selection and Assessment, 19*(1), 18-30.

Rubio, D., Berg-Weger, M., & Tebb, S. S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94-104.

Sackett, P. (1987). Assessment centers and content validity: some neglected issues. *Personnel Psychology, 40*(1), 13-25. doi: 10.1111/j.1744-6570.1987.tb02374.x

Sproule, C. F., (2009). *Rationale and research evidence supporting the use of content validation in personnel assessment.* A monograph of the International Personnel Assessment Council, p. 1-45.

Uebersax, J. S. (1988). Validity inferences from interobserver agreement. *Psychological Bulletin*, 1988, 104, 405-416.

Uebersax, J. S. (1993). Statistical Modelling of Expert Ratings on Medical Treatment Appropriateness. *Journal of the American Statistical Association, 88*(422), p. 421-427.

Uebersax, J. S. (2000). Agreement on Interval-Level Ratings. Retrieved November 11, 2012, from http://www.john-uebersax.com/stat/cont.htm

Van De Vijver, F., Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue euroéenne de psychologie appliquée, 54,* 119-135. doi:10.1016/j.erap.2003.12.004

Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education, 3*(1), 25-27.

UNIVERSITY OF JOHANNESBURG