

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

A Comparison of Task-Specific and Dimension-Specific Assessment Centres

Duncan J. R. Jackson

Members of the Supervisory Panel

Dr. Stephen G. Atkins (Chair)
Dr. Jennifer A. Stillman
Dr. Douglas Paton
Dr. Phillip E. Lowry

The real voyage of discovery consists not in seeking new landscapes, but in having new eyes
-Marcel Proust

To Whom It May Concern:

This is to state that, with respect to the research conducted for the Doctoral thesis entitled "A Comparison of Task-Specific and Dimension-Specific Assessment Centres" carried out by Duncan John Ross Jackson, the following statements are true:

- i) Reference to work other than that of the candidate has been appropriately acknowledged.
- ii) The research practice and ethical policies approved by Massey University have been complied with.
- iii) Although the current thesis guidelines request a word limit of 100,000, the current thesis was substantially completed prior to the introduction of this limit. (It consists of approximately 117,000 words.)

Stephen G. Atkins
Supervisor

 Date 23 Sept 2003

D.J.R. Jackson
Candidate



Date 23 Sept 2003

To Whom It May Concern:

This is to state that the research carried out for my Doctoral thesis entitled “*A Comparison of Task-Specific and Dimension-Specific Assessment Centres*” in the School of Psychology, Massey University, Albany Campus, New Zealand, is all my own work.

This is also to certify that the thesis material has not been used for any other degree.

D.J.R. Jackson
Candidate:



Date: 23 Sept 2003

To Whom It May Concern:

This is to state that the research conducted for the Doctoral thesis entitled “*A Comparison of Task-Specific and Dimension-Specific Assessment Centres*” was carried out by Duncan John Ross Jackson in accordance with the University’s Doctoral regulations.



Stephen G. Atkins
Supervisor

Date: 23 Sept 2003

This work is dedicated to my late grandfather, Mr. Ernest F. M. Wilson, who passed from this world at 6am on Sunday the 9th of February 2003. You were such a great and noble man, and your kindness, knowledge, wisdom, and humour will be so dearly missed. I wish I could have shared the contents of my dissertation with you, as I know you would have been keenly interested. You were one of the few people, in my younger years, who tempted me into the realisation that learning could be enjoyable. You shared your extensive knowledge of astronomy with me, and stirred a fascination, which remains today. Just prior to your passing, you said to me in your profound way; "You're my best friend". You are also my best friend, my dear grandfather. May you rest well, until we meet again.

Acknowledgements

Life-consuming ventures, such as the present Ph.D., are never performed on one's own, and due consideration must be given to all those who assisted me through this journey of discovery. From the early stages, I would like to thank Dr. Jennifer A. Stillman, A/Prof. Kerry Chamberlain, and Dr. Douglas Paton for assisting me to consolidate and formulate the methodology that would be used in my research. During the design of the assessment centres I used, I am indebted to Dr. Stephen G. Atkins, Dr. Felix E. Lopez and Dr. Phillip E. Lowry for their invaluable advice. For allowing me to gain access to organisations that use assessment centre methodology, I am grateful to Mr. Andrew Hambleton, Mr. Michael Hope, Ms. Helen Gribble, Ms. Rochelle McKay, Mr. Jason Clarke, Ms. Raewyn Bennett, Sqn Ldr Wanda Morris, Sqn Ldr Paul Gallagher, Sqn Ldr Laura Gillen, and Sqn Ldr Emma Davis. On the measures used in my research, I am grateful to Dr. Richard K. Wagner, Dr. Robert J. Sternberg, Dr. Albert Bandura, and Ms. Rebecca Tovey for their generosity. For the data analysis phase, I am indebted to Dr. Jennifer A. Stillman, Dr. Stephen G. Atkins, Dr. Robert L. Brennan, Dr. Richard J. Shavelson, Dr. Noreen M. Webb, Dr. George A. Marcoulides, Dr. John Spicer, Dr. Paul Barrett, Dr. Richard Fletcher, and Dr. John Hattie for their generous advice. For the laborious task of proofing, I am particularly indebted to Dr. Jennifer A. Stillman and Dr. Stephen G. Atkins. Thank you for your kindness, patience, and conscientiousness. In the study of assessment centres, thank you to Dr. Phillip E. Lowry, Dr. Peter Herriot, Dr. Ivan T. Robertson, and Dr. William A. Gorham for not taking the status quo at face value, and having the tenacity to stand against the prevailing view. Thank you to Dr. Filip Lievens for being at the forefront of contemporary assessment centre

research. Many thanks to Dr. Nikolaos Kazantzis for updating me on acceptable practices of assessment in clinical psychology. My gratefulness is extended to the members of my immediate family, my mother Mdm Annette M. Jackson, my father Mr. Michael J. R. Jackson, my two older brothers Mr. Hamish J. R. Jackson and Mr. Alistair J. R. Jackson and my younger sister, Ms. Daisy L. Jackson.

Thank you for believing in and supporting me throughout this time. Thanks to all my wonderful friends who have supported me through this journey, particularly Ms. Stella Cho, Mr. Victor Ng, Mr. Peter Johnston, and Mr. Shane Rowe. Lastly, I am forever indebted to all the members of my supervisory panel. Thank you for being both my friends and mentors. Your guidance has helped me to open my mind to the endless possibilities that could result from the study of psychology. The research contained, herein, was approved by the Massey University Albany Campus Human Ethics Committee, MUAHEC 00/047.

Abstract

Three studies were employed to further an understanding of a measurement quandary concerning assessment centres (ACs). A common theme associated with ACs is that they do not appear to measure the trait-based variables that they purport to. To compound this mystery, ACs are found to be predictive of outcome criteria; particularly criteria related to promotion. All three studies took varying perspectives on this measurement dilemma. The first study looked at particular traits that were not formally assessed in ACs, and whether these traits explained variance in overall AC ratings. No definitive evidence was found for this notion; however, tacit knowledge appeared to be associated with a small amount of variance in overall AC ratings in one of the samples under scrutiny. The second study looked at the extent to which assessors and candidates understood the models they were assessing and were being assessed under. Neither party appeared to distinguish trait-based, task-based, or other models as being more or less appropriate. While the first and second studies acknowledged some peripheral issues in the AC literature, the third study addressed the fundamental research question. Specifically, the third study investigated whether an alternative to the prevailing trait paradigm was needed. This study compared two models of assessment in a repeated measures design. One model treated the AC data as though they comprised situationally specific behavioural samples. The second model treated the data as though they were indicative of trait-based responses. Using a generalizability study, both models demonstrated similar psychometric characteristics, although only data treated under the situationally specific model held a conceptual justification. These findings suggest that the situationally specific task-based model presents a more appropriate means by which to treat AC ratings.

A Comparison of Task-Specific and Dimension-Specific Assessment Centres

Acknowledgements	i
Abstract	iii
List of Tables	5
List of Figures	7
Principal Notational Conventions	8
<u>Chapter One: Background and Hypotheses</u>	9
Background and History of the Assessment Centre Process	9
Group Exercises	15
Individual Exercises	15
Written Exercises	16
The Trait Paradigm	17
Construct Validity and the Exercise Effect	18
The Importance of Construct Validation in ACs	21
Construct Validation of ACs through the Nomological Network	24
Factors that May Improve AC Construct Validity: The Limited Information-Processing Model	25
Rating Dimensions Subsequent to Agreeing Upon Dimensional Ratings	27
Having Assessors Rate a Singular Dimension Across Exercises	29
Reducing Cognitive Load On Assessors and Organising Ratings	30
The Use of Video Recordings	35
Dimensional Transparency	35
Exercise Transparency and Opportunities To Express Behaviour	37
Form and Content of AC exercises	38
Factors that May Improve AC Construct Validity: The Expert Assessor Model	39
Frame of Reference Training	40
Employing Psychologists as Assessors	44
Attributing Variance to both Exercise and Dimensional Features	51
Overall Assessment Rating Integration Discussions in ACs	52
The Measurement of Latent Constructs in ACs	53
The Actual Criterion Contamination Explanation	54
The Subtle Criterion Contamination Explanation	56
The Self-Fulfilling Prophecy/Self-Efficacy Explanation	58
The Managerial Intelligence Explanation	61
The Impression Management Skill Explanation	65
Intelligence, Personality, and their Relationships with Overall Assessment Ratings (OARs)	70
The Behavioural and Interactionist Paradigms	72
The Performance Consistency Explanation	75
Evidence in Favour of a Task-Specific Approach	82

Summary	87
Overall Research Aim	88
Hypotheses	90
Study One, Hypothesis One	90
Study Two, Additional Research Question	90
Study Three, Hypothesis Two	90
<u>Chapter Two: Study One, Latent Trait Measurement in ACs</u>	92
Method	92
Prelude to Studies One and Two	92
Military Sample	93
Participants	93
Assessors	93
The RNZAF Selection Board	95
Selection Board Dimensions	95
Selection Board Exercises	97
Measures	99
Procedure	106
Organisational Sample	106
Participants	106
Assessors	108
The AC	108
AC Dimensions	109
AC Exercises	112
Results	113
Military Sample	114
Set One	115
Set Two	121
Supplementary Analysis for Set Two of the Military Sample	126
Organisational Sample	131
Discussion	137
Military Sample	138
Organisational Sample	139
Considerations	140
Analytical Limitations	141
Theoretical Implications	142
<u>Chapter Three: Study Two, Perceptions of Assessors and Candidates with respect to Measurement Models</u>	145
Method	145
Candidates	145
Assessors	145
Measures: Candidates	147
Measures: Assessors	148

Results	150
Candidates	150
Assessors	153
Discussion	157
Candidates	158
Assessors	158
Considerations	160
Theoretical Implications	161
<u>Chapter Four: Study Three, A Comparison of Task-Specific and Dimension-Specific ACs</u>	163
Method	163
Participants	163
Assessors	163
The AC	165
Task Analysis	166
TTA (Threshold Traits Analysis)	170
TTA Respondents	171
Summarising/Scoring Responses to the TTA	172
Presentation to the Managerial Level SME Panel	174
Classification and Extrapolation of Tasks into Dimensions	175
AC Task Ratings and Dimensions	176
AC Exercises	177
Evaluation Approach	178
Assessor Training and the Assessment Procedure	178
Procedure	181
Results	182
Generalizability Study	184
Factor Analysis	192
Varimax Rotation	193
Direct Oblimin Rotation	196
Confirmatory Factor Analysis	198
Discussion	207
Generalizability Study	207
Factor Analysis	212
Confirmatory Factor Analysis	214
Considerations	215
Theoretical Implications	218
<u>Chapter Five: General Discussion</u>	222
References	234
<u>Appendix I: Pilot for Study Three</u>	251

Method	251
Results and Discussion	262
<u>Appendix II: Introduction to Generalizability Theory</u>	273
<u>Appendix III: Abridged Assessment Centre Manual and Training Guide For Farmers Merchandiser, General Sales and One On One Sales Roles: Including General Sales Exercises</u>	300

List of Tables

Table Number	Title	Page Number
1	<i>Average Convergent and Discriminant Validity Coefficients of Assessor Ratings in a Sample of AC Studies</i>	22
2	<i>Predictive Validity of Various Designs of AC Research</i>	56
3	<i>Various Criteria Used and their Predictive Validity with AC Outcomes</i>	58
4	<i>Demographic Statistics, Candidates, Study One Military Sample</i>	94
5	<i>Demographic Statistics, Assessors, Study One Military Sample</i>	95
6	<i>Demographic Statistics, Candidates, Study One Organisational Sample</i>	107
7	<i>Overall Means and Standard Deviations for Measures Employed in Set One of the Military Sample</i>	116
8	<i>Bivariate Correlations Between Measures Employed in Set One of the Military Sample</i>	117
9	<i>Multiple Regression Analysis for the Prediction of OARs in Set One of the Military Sample</i>	118
10	<i>Spearman's Rho Between Measures Employed in Set One of the Military Sample</i>	120
11	<i>Overall Means and Standard Deviations for Measures Employed in Set Two of the Military sample</i>	121
12	<i>Bivariate Correlations Between Measures Employed in Set Two of the Military Sample</i>	122
13	<i>Multiple Regression Analysis for the Prediction of OARs in Set Two of the Military Sample</i>	124
14	<i>Spearman's Rho Between Measures Employed in Set Two of the Military Sample</i>	125
15	<i>Bivariate Correlations Between Measures Employed in Supplementary Set Two of the Military Sample</i>	128
16	<i>Multiple Regression Analysis for the Prediction of OARs in Supplementary Set Two of the Military Sample</i>	129
17	<i>Spearman's Rho Between Measures Employed in Supplementary Set Two of the Military Sample</i>	130
18	<i>Overall Means and Standard Deviations for Measures Employed in the Organisational Sample</i>	133
19	<i>Bivariate Correlations Between Measures Employed in the Organisational Sample</i>	133
20	<i>Multiple Regression Analysis for the Prediction of OARs in the Organisational Sample</i>	135
21	<i>Spearman's Rho Between Measures Employed in the Organisational Sample</i>	136
22	<i>Demographic Statistics, Candidates, Study Two Military Sample</i>	146
23	<i>Demographic Statistics, Assessors, Study Two Military Sample</i>	147

24	<i>Model Assumed to Underlie Assessment: Candidates</i>	151
25	<i>Model Assumed to Guide Assessment: Candidates</i>	152
26	<i>Model Assumed to Guide Assessment After Being Informed of the Measurement Problems in ACs: Candidates</i>	153
27	<i>Usefulness of Individual Dimensions: Assessors</i>	154
28	<i>Dimensions Perceived as Being Seen Exhibited Across All Exercises: Assessors</i>	155
29	<i>Model Assumed to Guide Assessment: Assessors</i>	156
30	<i>Model Assumed to Guide Assessment After Being Informed Of The Measurement Problems In ACs: Assessors</i>	157
31	<i>Demographic Statistics, Candidates, Study Three Private Sector Sample</i>	164
32	<i>Grand Means and SDs of the Behavioural Ratings (Within Exercises) in the Task-Specific AC</i>	183
33	<i>Grand Means and SDs of the Dimension Ratings (Across Exercises) in the Dimension Specific AC</i>	183
34	<i>Generalizability Study Comparing a Task-Specific with a Dimension-Specific AC in a Repeated Measures Design for the Organisational Sample</i>	187
35	<i>Relative and Absolute Error, Generalizability and Phi Coefficients and Interrater Reliability for the Balanced Task-Specific AC</i>	190
36	<i>Relative and Absolute Error, Generalizability and Phi Coefficients and Interrater Reliability for the Dimension-Specific AC</i>	191
37	<i>Generalizability Study Showing the Results of the Unbalanced Task-Specific AC for the Organisational Sample</i>	192
38	<i>Rotated Factor Matrix for the Task-Specific AC Ratings</i>	194
39	<i>Rotated Factor Matrix for the Dimension-Specific AC Ratings</i>	195
40	<i>Rotated Pattern Matrix for the Task-Specific AC Ratings</i>	197
41	<i>Rotated Pattern Matrix for the Dimension-Specific AC Ratings</i>	198
42	<i>Standardised Factor Loadings for Model One: The Abridged Task-Specific CFA Model</i>	201
43	<i>Selected Goodness-Of-Fit Indices for Model One: The Abridged Task-Specific CFA Model</i>	202
44	<i>Standardised Factor Loadings for Model Two: The Dimension-Specific CFA Model</i>	203
45	<i>Selected Goodness-Of-Fit Indices for Model Two: The Dimension-Specific CFA Model</i>	204
46	<i>Standardised Factor Loadings for Model Three: The Exercise Effect CFA Model</i>	205
47	<i>Selected Goodness-Of-Fit Indices for Model Three: The Exercise Effect CFA Model</i>	206
48	<i>Advantages of the Task-Specific Approach Relative to the Dimension-Specific Approach to AC Design</i>	228

List of Figures

Figure Number	Caption	Page Number
1	<i>Competency/Exercise Matrix for Study One, Organisational Sample</i>	110
2	<i>Variance Components and Confidence Intervals for Each Effect and Interaction in the Task-Specific AC</i>	189
3	<i>Variance Components and Confidence Intervals for Each Effect and Interaction in the Dimension-Specific AC</i>	189
4	<i>Model One: Abridged Task-Specific CFA Model</i>	199
5	<i>Model Two: Dimension-Specific CFA Model</i>	202
6	<i>Model Three: The Exercise Effect CFA Model</i>	204

Principal Notational Conventions for Generalizability Studies

- p The main effect for persons, the object of measurement in G studies.
- x The main effect for assessment centre exercises. This and any other source of variance in a G study, except for the object of measurement, is termed a *facet*.
- d The main effect for dimensions, traits, or competencies. These constructs are assumed to have a quality that is relatively stable and enduring across assessment exercises. This and any other source of variance in a G study, except for the object of measurement, is termed a facet.
- px The interaction term for two (or more) facets in a G study.
- pxd,e The interaction between all the facets in a G study followed by an 'e' indicates the error term for the model. This is the component of variance that is attributable to undifferentiated error.
- $i:x$ The presence of a colon (:) indicates that one facet is nested within another. In this case, the facet 'i' (items) is nested within 'x' (exercises). This occurs in a task-specific assessment centre, because each exercise has its own associated set of items.
- σ_{Rel}^2 Relative error term. Used to calculate measurement error associated with all of the components of variance that compare the standing of individuals relative to one another. This term is used in the calculation of the G coefficient.
- σ_{Abs}^2 Absolute error term. Used to calculate measurement error associated with all of the components of variance that relate to absolute decisions. That is, decisions that have a cut-off point, or a pass/fail criterion. This term is used in the calculation of the Phi coefficient.
- $E\rho_{\text{Rel}}^2$ The G coefficient for relative decisions. This is presented on a scale from 0, indicating poor generalizability, to 1, indicating excellent generalizability.
- ϕ The Phi coefficient for absolute decisions. This is presented on a scale from 0, indicating poor generalizability, to 1, indicating excellent generalizability.