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.)

D.J.R. Jackson
Candidate

Date 23 Sept 2005

Stephen G. Atkins
Supervisor

Date 23 Sept 2005
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

Abstract

List of Tables

List of Figures

Principal Notational Conventions

Chapter One: Background and Hypotheses

Background and History of the Assessment Centre Process

Group Exercises

Individual Exercises

Written Exercises

The Trait Paradigm

Construct Validity and the Exercise Effect

The Importance of Construct Validation in ACs

Construct Validation of ACs through the Nomological Network

Factors that May Improve AC Construct Validity: The Limited Information-Processing Model

Rating Dimensions Subsequent to Agreeing Upon Dimensional Ratings

Having Assessors Rate a Singular Dimension Across Exercises

Reducing Cognitive Load On Assessors and Organising Ratings

The Use of Video Recordings

Dimensional Transparency

Exercise Transparency and Opportunities To Express Behaviour

Form and Content of AC exercises

Factors that May Improve AC Construct Validity: The Expert Assessor Model

Frame of Reference Training

Employing Psychologists as Assessors

Attributing Variance to both Exercise and Dimensional Features

Overall Assessment Rating Integration Discussions in ACs

The Measurement of Latent Constructs in ACs

The Actual Criterion Contamination Explanation

The Subtle Criterion Contamination Explanation

The Self-Fulfilling Prophecy/Self-Efficacy Explanation

The Managerial Intelligence Explanation

The Impression Management Skill Explanation

Intelligence, Personality, and their Relationships with Overall Assessment Ratings (OARs)

The Behavioural and Interactionist Paradigms

The Performance Consistency Explanation

Evidence in Favour of a Task-Specific Approach
Summary

Overall Research Aim

Hypotheses
- Study One, Hypothesis One
- Study Two, Additional Research Question
- Study Three, Hypothesis Two

Chapter Two: Study One, Latent Trait Measurement in ACs

Method
- Prelude to Studies One and Two
- Military Sample
  - Participants
  - Assessors
  - The RNZAF Selection Board
  - Selection Board Dimensions
  - Selection Board Exercises
  - Measures
  - Procedure
- Organisational Sample
  - Participants
  - Assessors
  - The AC
  - AC Dimensions
  - AC Exercises

Results
- Military Sample
  - Set One
  - Set Two
  - Supplementary Analysis for Set Two of the Military Sample
- Organisational Sample

Discussion
- Military Sample
- Organisational Sample
- Considerations
- Analytical Limitations
- Theoretical Implications

Chapter Three: Study Two, Perceptions of Assessors and Candidates with respect to Measurement Models

Method
- Candidates
- Assessors
- Measures: Candidates
- Measures: Assessors
Chapter Four: Study Three, A Comparison of Task-Specific and Dimension-Specific ACs

Method

Participants
Assessors
The AC
Task Analysis
TTA (Threshold Traits Analysis)
TTA Respondents
Summarising/Scoring Responses to the TTA
Presentation to the Managerial Level SME Panel
Classification and Extrapolation of Tasks into Dimensions
AC Task Ratings and Dimensions
AC Exercises
Evaluation Approach
Assessor Training and the Assessment Procedure
Procedure

Results

Generalizability Study
Factor Analysis
Varimax Rotation
Direct Oblimin Rotation
Confirmatory Factor Analysis

Discussion

Generalizability Study
Factor Analysis
Confirmatory Factor Analysis
Considerations
Theoretical Implications

Chapter Five: General Discussion

References

Appendix I: Pilot for Study Three
### List of Tables

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Title</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average Convergent and Discriminant Validity Coefficients of Assessor Ratings in a Sample of AC Studies</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Predictive Validity of Various Designs of AC Research</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>Various Criteria Used and their Predictive Validity with AC Outcomes</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>Demographic Statistics, Candidates, Study One Military Sample</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>Demographic Statistics, Assessors, Study One Military Sample</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>Demographic Statistics, Candidates, Study One Organisational Sample</td>
<td>107</td>
</tr>
<tr>
<td>7</td>
<td>Overall Means and Standard Deviations for Measures Employed in Set One of the Military Sample</td>
<td>116</td>
</tr>
<tr>
<td>8</td>
<td>Bivariate Correlations Between Measures Employed in Set One of the Military Sample</td>
<td>117</td>
</tr>
<tr>
<td>9</td>
<td>Multiple Regression Analysis for the Prediction of OARs in Set One of the Military Sample</td>
<td>118</td>
</tr>
<tr>
<td>10</td>
<td>Spearman's Rho Between Measures Employed in Set One of the Military Sample</td>
<td>120</td>
</tr>
<tr>
<td>11</td>
<td>Overall Means and Standard Deviations for Measures Employed in Set Two of the Military Sample</td>
<td>121</td>
</tr>
<tr>
<td>12</td>
<td>Bivariate Correlations Between Measures Employed in Set Two of the Military Sample</td>
<td>122</td>
</tr>
<tr>
<td>13</td>
<td>Multiple Regression Analysis for the Prediction of OARs in Set Two of the Military Sample</td>
<td>124</td>
</tr>
<tr>
<td>14</td>
<td>Spearman's Rho Between Measures Employed in Set Two of the Military Sample</td>
<td>125</td>
</tr>
<tr>
<td>15</td>
<td>Bivariate Correlations Between Measures Employed in Supplementary Set Two of the Military Sample</td>
<td>128</td>
</tr>
<tr>
<td>16</td>
<td>Multiple Regression Analysis for the Prediction of OARs in Supplementary Set Two of the Military Sample</td>
<td>129</td>
</tr>
<tr>
<td>17</td>
<td>Spearman's Rho Between Measures Employed in Supplementary Set Two of the Military Sample</td>
<td>130</td>
</tr>
<tr>
<td>18</td>
<td>Overall Means and Standard Deviations for Measures Employed in the Organisational Sample</td>
<td>133</td>
</tr>
<tr>
<td>19</td>
<td>Bivariate Correlations Between Measures Employed in the Organisational Sample</td>
<td>133</td>
</tr>
<tr>
<td>20</td>
<td>Multiple Regression Analysis for the Prediction of OARs in the Organisational Sample</td>
<td>135</td>
</tr>
<tr>
<td>21</td>
<td>Spearman's Rho Between Measures Employed in the Organisational Sample</td>
<td>136</td>
</tr>
<tr>
<td>22</td>
<td>Demographic Statistics, Candidates, Study Two Military Sample</td>
<td>146</td>
</tr>
<tr>
<td>23</td>
<td>Demographic Statistics, Assessors, Study Two Military Sample</td>
<td>147</td>
</tr>
</tbody>
</table>
24 Model Assumed to Underlie Assessment: Candidates
25 Model Assumed to Guide Assessment: Candidates
26 Model Assumed to Guide Assessment After Being Informed of the Measurement Problems in ACs: Candidates
27 Usefulness of Individual Dimensions: Assessors
28 Dimensions Perceived as Being Seen Exhibited Across All Exercises: Assessors
29 Model Assumed to Guide Assessment: Assessors
30 Model Assumed to Guide Assessment After Being Informed Of The Measurement Problems In ACs: Assessors
31 Demographic Statistics, Candidates, Study Three Private Sector Sample
32 Grand Means and SDs of the Behavioural Ratings (Within Exercises) in the Task-Specific AC
33 Grand Means and SDs of the Dimension Ratings (Across Exercises) in the Dimension Specific AC
34 Generalizability Study Comparing a Task-Specific with a Dimension-Specific AC in a Repeated Measures Design for the Organisational Sample
35 Relative and Absolute Error, Generalizability and Phi Coefficients and Interrater Reliability for the Balanced Task-Specific AC
36 Relative and Absolute Error, Generalizability and Phi Coefficients and Interrater Reliability for the Dimension-Specific AC
37 Generalizability Study Showing the Results of the Unbalanced Task-Specific AC for the Organisational Sample
38 Rotated Factor Matrix for the Task-Specific AC Ratings
39 Rotated Factor Matrix for the Dimension-Specific AC Ratings
40 Rotated Pattern Matrix for the Task-Specific AC Ratings
41 Rotated Pattern Matrix for the Dimension-Specific AC Ratings
42 Standardised Factor Loadings for Model One: The Abridged Task-Specific CFA Model
43 Selected Goodness-Of-Fit Indices for Model One: The Abridged Task-Specific CFA Model
44 Standardised Factor Loadings for Model Two: The Dimension-Specific CFA Model
45 Selected Goodness-Of-Fit Indices for Model Two: The Dimension-Specific CFA Model
46 Standardised Factor Loadings for Model Three: The Exercise Effect CFA Model
47 Selected Goodness-Of-Fit Indices for Model Three: The Exercise Effect CFA Model
48 Advantages of the Task-Specific Approach Relative to the Dimension-Specific Approach to AC Design
## List of Figures

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

\( \sigma^2_{\text{Rel}} \) 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.

\( \sigma^2_{\text{Abs}} \) 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.

\( E_{\text{Rel}}^2 \) The G coefficient for relative decisions. This is presented on a scale from 0, indicating poor generalizability, to 1, indicating excellent generalizability.

\( \phi \) The Phi coefficient for absolute decisions. This is presented on a scale from 0, indicating poor generalizability, to 1, indicating excellent generalizability.