REFERENCE REPORTS:

A META-ANALYTIC REVIEW OF PREDICTIVE VALIDITY AND AN EXPERIMENTAL STUDY OF RATING ACCURACY

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Psychology at Massey University

KARL BRANDON PAJO

1996
Massey University Library
Thesis-Copyright Form

Title of Thesis:

(1) (a) I give permission for my thesis to be made available to readers in Massey University Library under conditions determined by the Librarian.

(b) I do not wish my thesis to be made available to readers without my written consent for ... months.

(2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

(b) I do not wish my thesis, or a copy, to be sent to another institution without my written consent for ... months.

(3) (a) I agree that my thesis may be copied for Library use.

(b) I do not wish my thesis to be copied for Library use for ... months.

Signed ____________________________

Date 31/5/96

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME AND ADDRESS

DATE
Abstract

Reference reports are a commonly used selection method in New Zealand and overseas. Although popular with practitioners, they have attracted little attention from researchers. To ascertain the predictive validity of reference reports a meta-analytic review was conducted. Results of the preliminary analysis provided a bare-bones estimated mean validity of 0.15, and a fully corrected estimated mean validity of 0.32. Substantial variance remained unaccounted for following corrections for sampling error. Subsequent moderated meta-analyses, based on degree of structure and psychometric soundness of the reference reports, was found to account for much of the variation in observed validity coefficients. Highly structured reports were found to be consistently superior to unstructured reports. Improvements in the validity of highly structured reports can be attributed to the control of leniency in ratings. However, no studies to date have evaluated the accuracy of referees' ratings. Drawing on the performance rating literature, an experimental study examining the influence of scale format, ratee characteristics, rating purpose, and rater affect on the accuracy of ratings was implemented. Asymmetrical, positively toned scales were found to reduce leniency in ratings compared to unstructured and Likert-type rating forms. Raters who expressed liking for the ratee were more lenient in their ratings compared to raters who expressed neutral or antagonistic feelings toward the ratee. No significant effects for rating purpose and ratee characteristics were apparent. Overall, the investigation points to deficiencies in the way reference reports are presently employed, and highlights the need for a more rigorous approach in their development and application.
Acknowledgments

I am indebted to my supervisors, John Podd and Mike Smith, for the support and assistance they offered during the research and writing of this dissertation. The common sense and sound advice they have provided has been much appreciated. I would like to express my special thanks to John, who willingly took on the major supervisory duties when Mike left Massey University to take up another position.

I am very grateful for the encouragement and aid forthcoming from my colleagues in the Departments of Psychology, and Human Resource Management. Not only are they a great bunch of people to work with, but they have proved to be good friends as well. My thanks also go to John Spicer who helped clarify some of the statistical issues that arose throughout the course of the research.

I would also like to express my gratitude to my friends and my family who have been a supportive and understanding crew, especially on those occasions when I couldn’t “come out to play.”

Finally, it is with deep appreciation that I wish to acknowledge the contribution of my partner, Jacqui. I am very grateful for her constant encouragement, patience, and support during the long nights and weekends when she would have much preferred to see me at home. Thank you.
TABLE OF CONTENTS

Abstract ......................................................................................................................... ii

Acknowledgements ........................................................................................................ iii

Chapter 1
   Overview .................................................................................................................... 1

Chapter 2
   Literature Review: Reference Reports ...................................................................... 5
      Surveys of Use ........................................................................................................... 6
      Content of Reference Reports ................................................................................. 11
      Psychometric Issues ............................................................................................... 15
         Reliability ............................................................................................................. 15
         Validity ................................................................................................................ 20
            Source of the Recommendation ....................................................................... 20
            Target Population ............................................................................................. 22
            Format and Content .......................................................................................... 23
            Miscellaneous Threats to Validity ..................................................................... 27
            Meta-Analytic Studies ....................................................................................... 31
   The Meta-Analysis ...................................................................................................... 32

Chapter 3
   Meta-Analysis .......................................................................................................... 34
      Method ..................................................................................................................... 34
         Literature Search .................................................................................................. 34
         Decision Rules for Coding Studies into the Meta-Analysis .................................. 35
<table>
<thead>
<tr>
<th>Meta-Analysis Method Used</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>43</td>
</tr>
<tr>
<td>Post hoc Analyses</td>
<td>45</td>
</tr>
<tr>
<td>Discussion</td>
<td>51</td>
</tr>
</tbody>
</table>

**Chapter 4**

**Literature Review: Performance Rating** | 60 |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater Training</td>
<td>60</td>
</tr>
<tr>
<td>Cognitive Processes in Performance Rating</td>
<td>64</td>
</tr>
<tr>
<td>Rating Scale Format</td>
<td>65</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>69</td>
</tr>
<tr>
<td>Rater Acquaintance and Affect</td>
<td>71</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>79</td>
</tr>
<tr>
<td>Ratee Characteristics</td>
<td>79</td>
</tr>
<tr>
<td>Race</td>
<td>80</td>
</tr>
<tr>
<td>Sex</td>
<td>82</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>84</td>
</tr>
<tr>
<td>Purpose of Rating</td>
<td>86</td>
</tr>
<tr>
<td><strong>Rater Motivation</strong></td>
<td>92</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>95</td>
</tr>
<tr>
<td>Rater Accuracy</td>
<td>96</td>
</tr>
</tbody>
</table>

**Chapter 5**

**Experimental Study** | 99 |

**Chapter 6**

**Method** | 103 |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedural Overview</td>
<td>103</td>
</tr>
<tr>
<td>Participants</td>
<td>103</td>
</tr>
<tr>
<td>Procedure</td>
<td>104</td>
</tr>
</tbody>
</table>
Experimental Methods .......................................................... 105
Vignette ...................................................................................... 105
Likability Ratings ................................................................. 108
Demographic Questionnaire ..................................................... 108
Rating Forms ............................................................................ 109
Leniency Scale ........................................................................ 110
True Scores ................................................................................ 112
Dependent Variables ................................................................ 114
Accuracy Measures ................................................................. 114
Error Measures ........................................................................ 116
Data Analysis ............................................................................ 117
Rescaling of Ratings and Measures .......................................... 117
Analyses .................................................................................... 119

Chapter 7

Results ...................................................................................... 121
Demographic Questionnaire ..................................................... 121
Tests of Hypotheses ................................................................ 127
Accuracy Measures ................................................................ 131
Error Measures ........................................................................ 134
Comparisons Between Measures ............................................. 136
Regression Analysis of Error Scores ......................................... 138
Prediction of Leniency ............................................................... 140

Chapter 8

Discussion ................................................................................ 147
The Rating Instrument .............................................................. 148
Rater Affect .............................................................................. 151
Rating Purpose .......................................................................... 157
Ratee Characteristics .............................................................. 161
<table>
<thead>
<tr>
<th>Criterion Measures</th>
<th>...............................................................................................................</th>
<th>163</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey of Reference Reports</td>
<td>...............................................................................................................</td>
<td>165</td>
</tr>
<tr>
<td>Limitations of the Current Study</td>
<td>...............................................................................................................</td>
<td>166</td>
</tr>
<tr>
<td>Future Research</td>
<td>...............................................................................................................</td>
<td>170</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>...............................................................................................................</td>
<td>175</td>
</tr>
</tbody>
</table>

**References** ............................................................................................................... 179

**APPENDIX 1**

Letter Requesting Participation in the Study ..................................................................... 205

**APPENDIX 2**

Teaching Vignette and Likability Scale .............................................................................. 208

**APPENDIX 3**

Demographic Questionnaire ............................................................................................... 217

**APPENDIX 4**

Unstructured Rating Forms Developed for the Referee's Report and for the Performance Appraisal ........................................................................................................... 219

**APPENDIX 5**

Likert-Type Rating Forms Developed for the Referee's Report and for the Performance Appraisal ........................................................................................................... 222

**APPENDIX 6**

Asymmetrical Rating Forms Developed for the Referee's Report and for the Performance Appraisal ........................................................................................................... 227
APPENDIX 7

Gender-Typed versions of the Leniency Scale ........................................... 230

APPENDIX 8

Comparison of Rating Form Accuracy Using the Original Rating Metric ................................................................. 233
# LIST OF TABLES

**TABLE 1**  
*Studies, sample type, total number of participants and validity coefficients contributed by each study to the meta-analysis*  
Page 35

**TABLE 2**  
*Meta-analysis of the predictive validity of reference reports*  
Page 44

**TABLE 3**  
*Meta-analysis of the predictive validity of reference reports moderated by structure*  
Page 48

**TABLE 4**  
*Meta-analysis of the predictive validity of reference reports moderated by structure and criterion type*  
Page 49

**TABLE 5**  
*Mean reliability (average correlations) of judges’ ratings for responses on the unstructured rating form*  
Page 117

**TABLE 6**  
*Summary of responses to the yes/no items in the demographic questionnaire*  
Page 122

**TABLE 7**  
*Descriptive statistics for questions from the demographic questionnaire*  
Page 122

**TABLE 8**  
*Significant main effects for type of form from the overall ANOVAs calculated for each teaching dimension*  
Page 128
TABLE 9
Means and standard deviations of ratings of teaching dimensions from three different rating forms. 129

TABLE 10
Results from t-test comparing mean ratings from three different forms. 130

TABLE 11
Mean accuracy values as a function of purpose and rating form. 131

TABLE 12
Results of t-tests comparing mean accuracy values for ratings from three different forms. 132

TABLE 13
Mean error values for ratings from three different forms used for performance appraisal and reference report purposes. 135

TABLE 14
Correlations among rater error measures. 136

TABLE 15
Correlations among rater accuracy measures. 137

TABLE 16
Correlations between accuracy and error measures. 138
TABLE 17
Results from a standard multiple regression analysis using four error measures to predict each of four accuracy scores .......................... 139

TABLE 18
Correlations among variables included in the standard regression analysis for the prediction of leniency in ratings .............................. 141

TABLE 19
Results from a standard multiple regression using six personal and contextual variables to predict leniency in ratings ............................ 143

TABLE 20
Results from the hierarchical multiple regression analysis for the prediction of leniency in ratings after controlling for the effects of rating form (Analysis 1) and for the effects of rating form and Likability (Analysis 2) .................................................. 145

TABLE 21
Referees' access to information about task and interpersonal behaviours and results ............................................................................. 174

TABLE 22
Results of t-tests comparing mean accuracy values for ratings from three different forms using the original rating metric .......................... 234
LIST OF FIGURES

FIGURE 1  
Distribution of validity coefficients from studies in the meta-analysis ................................................................. 39

FIGURE 2  
Rated usefulness of reference reports ........................................................................................................... 124