The Validity of Assessment Centres for the Prediction of Supervisory Performance Ratings: A meta-analysis

Eran Hermelin*, Filip Lievens** and Ivan T. Robertson***

*Psychology Group, Manchester Business School, University of Manchester, Manchester M15 6PB, UK. info@eranhermelin.com
**Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
***Robertson Cooper Ltd., Manchester, UK

The current meta-analysis of the selection validity of assessment centres aims to update an earlier meta-analysis of assessment centre validity. To this end, we retrieved 26 studies and 27 validity coefficients ($N = 5850$) relating the Overall Assessment Rating (OAR) to supervisory performance ratings. The current study obtained a corrected correlation of .28 between the OAR and supervisory job performance ratings (95% confidence interval $.24 \leq \rho \leq .32$). It is further suggested that this validity estimate is likely to be conservative given that assessment centre validities tend to be affected by indirect range restriction.

1. Introduction

In human resources management, assessment centres essentially serve two purposes (Thornton, 1992). Their first and most traditional purpose is selection and promotion. In these assessment centre applications, the so-called Overall Assessment Rating (OAR) plays a predominant role as selection and promotion decisions are contingent upon it. Gaugler, Rosenthal, Thornton, and Bentson (1987) reported findings supporting the validity of predictions made on the basis of the OAR. Specifically, their meta-analysis estimated the mean operational validity of the OAR to be .36 with respect to the criterion of job performance ratings. As regards to their second purpose, assessment centres are increasingly used for developing managerial talent. In developmental assessment centres, the focus shifts from the OAR to assessment centre dimensions which serve as the basis for providing participants with detailed feedback about their strengths and weaknesses. Recently, Arthur, Day, McNelly, and Edens (2003) reported evidence supporting the validity of predictions made on the basis of assessment centre dimensions. In particular, their meta-analysis focused on six main assessment centre dimensions (consideration/awareness of others, communication, drive, influencing others, organizing and planning, and problem solving), with validities varying from .25 to .39.

Thus, although a recent meta-analysis updated the validity of predictions made on the basis of assessment centre dimensions (Arthur *et al.*, 2003), the results of

This paper is based in part on the Doctoral thesis submitted by the first author to the University of Manchester Institute of Science and Technology in fulfillment of the requirements for the degree of PhD in Management Science.

9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main St., Malden, MA, 02148, USA

Gaugler *et al*. (1987) still serve as the 'gold standard' of assessment centre validity estimation on the basis of the OAR. However, since 1987, new validity studies have been conducted that were obviously not included in the Gaugler *et al*. (1987) meta-analysis. Accordingly, this study meta-analysed studies that were not included in the Gaugler *et al*. (1987) meta-analysis (from 1985 onward). In keeping with the Gaugler *et al*. study, we focus on the selection validity of assessment centres (i.e., their ability to select the best candidates for a given job) instead of on the validity of assessment centre dimensions (see Arthur *et al*., 2003). Supervisory performance ratings served as the criterion measure in the current meta-analysis.

2. Method

2.1. Database

We used a number of strategies to identify validity studies potentially suited for inclusion in the current meta-analysis. First, a computerized search of various electronic databases was conducted (PsycInfo, Social Sciences Citation Index, etc.). Second, a computerized search of the British Psychological Society database of UK-based Chartered Psychologists was undertaken and academics and practitioners were contacted to identify individuals who may have access to unpublished assessment centre validity data. Third, around 20 of the top companies in the FTSE 500 index and four of the United Kingdom's largest occupational psychology firms were contacted.

2.2. Inclusion and coding criteria

We scrutinized the studies retrieved and included them in our final database if they met the following four inclusion criteria. First, studies were considered if they referred to an 'assessment centre'. An assessment centre was defined on the basis of the following criteria: (a) The selection procedure included two or more selection methods, at least one of which was an individual/group simulation exercise, (b) one of more assessors were required to directly observe assessees' behaviour in at least one of the simulation exercises, (c) evaluation of assessees' performance on the selection methods included in the selection procedure was (or could be) integrated into an OAR by a clinical or statistical integration procedure, or both; (d) the selection procedure lasted for at least 2 hours.

As we wanted to update the prior meta-analysis of the selection validity of assessment centres, a second criterion specified that only studies published or completed from 1985 onwards were considered for inclusion in the meta-analytic database.

Third, we used Borman's (1991) definition of supervisory job performance ratings, which he defined as 'an estimate of individuals' performance made by a supervisor' (p. 280). This estimate could be either an overall or a multi-dimensional performance evaluation. Hence, ratings of potential, objective measures of performance, performance tests, and ratings made by peers were excluded.

Fourth, studies had to provide sufficient information to be coded. As most of the studies did not report all the necessary information, the first author attempted to contact the authors of these studies.

On the basis of these four inclusion criteria 26 studies with 27 non-overlapping validity coefficients were included in the meta-analysis. The total N was 5850 (as compared with $N = 4180$ of Gaugler *et al*., 1987). Of these studies, 23 had been published, one was presented at an international conference, and two were unpublished. The earliest study included was published in 1985, whereas the most recent study was conducted in 2005.

The coding of the 27 validity coefficients which constituted the final meta-analytic dataset was conducted separately by the first and second authors. On the basis of a sample of studies coded by the authors, a reliability check revealed that when both authors entered a coding, their coding agreed in 85% of cases. The full coding scheme is available from the first author. At the end of this procedure, the separately coded datasets were compared and any disagreements were resolved between the two authors.

2.3. Evaluation of publication bias

The data were subjected to two additional examinations. First, to explore the possibility that assessment centre validities were somehow related to extraneous factors which could not be regarded as potential moderators, corrected validities were correlated with the study completion/publication date and with the time interval between the collection of OARs and supervisory performance ratings. The values of these two correlations were $r = -.05$ and .10, respectively. Hence, there was no consistent relationship between either of these two factors and assessment centre validities.

Second, a funnel plot was performed to check the distribution of corrected validities (see Egger, Smith, Schneider, & Minder, 1997). The idea behind this procedure is to plot corrected validities against sample size, so as to examine whether sampled validities appear to be free of publication bias. As the degree of sampling error depends on sample size it is to be expected that the spread of validities would become progressively smaller as sample sizes increased, thereby creating a scatter plot resembling a symmetrical inverted funnel. Should the sampling be biased, the funnel plot would be asymmetrical (Egger *et al*., 1997). For the studies included in this meta-analysis, the scatter plot of validities resembled an inverted funnel with validity coefficients based on small samples showing considerable variation, whereas those based on larger samples tended to converge on the mean meta-analytic validity coefficient. There was however a tendency for validities not to be evenly distributed around the mean, with six coefficients located under the meta-analytic mean, one coefficient corresponding to the meta-analytic mean, and 20 coefficients located above the meta-analytic mean. There was a tendency for studies with larger sample sizes to be more evenly distributed around the meta-analytic mean than studies based on smaller sample sizes. The study contributing the largest sample size to the meta-analytic dataset (28% of total cases) was positioned in the middle of the distribution of validities and so did not skew the outcome of the meta-analysis.

2.4. Meta-analytic procedure

The dataset was analysed by following the Hunter and Schmidt (1990) procedures for individually correcting correlations for experimental artifacts. In instances in which these were deemed not to be sufficiently detailed for the purposes of the current study, advice was solicited directly from the book's second author (F. Schmidt, personal communication, 2002). We were able to obtain range restriction data for 20 out of the 27 validity coefficients included in the dataset. These 20 coefficients were hence individually coded for range restriction. In the absence of specific information about the range restriction ratios of the remaining seven coefficients, they were assigned the mean of the range restriction ratios coded for the 20 coefficients individually coded for range restriction (see the Appendix A).

As reliabilities for supervisory performance ratings were typically not mentioned in the studies included in our meta-analytic dataset, we decided to use the best available reliability estimates for supervisory performance ratings. In fact, two large scale meta-analyses found .52 to be the average criterion reliability estimate for supervisory performance ratings (Salgado *et al*. 2003; Viswesvaran, Ones, & Schmidt, 1996). Hence, we decided to use the value of .52 as the criterion reliability estimate for all 27 validity coefficients.

Although there now exist procedures to correct for indirect range restriction (Hunter, Schmidt, & Le, 2006; Schmidt, Oh, & Le, 2006), we were not able to perform this correction as indirect range restriction data were not available in the primary studies. We were therefore unable to go beyond the standard practice of correcting the magnitude of observed validities for the presence of direct range restriction and criterion unreliability.

3. Results

3.1. Assessment centre validity

As shown in Table 1, the mean observed *r* based on a total sample size of 5850 was .17. Correcting this coefficient for direct range restriction in the predictor variable increased its value to .20. When the coefficient was also corrected for criterion unreliability, the population estimate for the correlation between OARs and supervisory performance ratings, increased to .28 [95% confidence interval (CI) $= .24 \leq \rho \leq .32$]. Details of the distribution of artifacts used to individually correct observed validity coefficients, are provided in Table 1, which shows that 84% of variance in validity coefficients may be explicable in terms of sampling error. Consequently, once the variance theoretically contributed by sampling error was removed, little unexplained variance remained. Thus, the detection of potential moderator variables was unlikely. Nevertheless, we tested for various moderators (e.g., number of dimensions assessed, number of different selection methods, type of integration procedure used). As could be expected, none of these moderators was significant.

4. Discussion

The typically used meta-analytic estimate of the validity of assessment centre OARs (Gaugler *et al*., 1987) is based on studies conducted prior to 1985, some of which are now over 50 years old. However, in the last 20 years, many new assessment centre validation studies have been conducted. Although Arthur *et al*. (2003) recently provided an updated estimate of the validity of assessment centre *dimensions*, it is also important to provide an updated estimate of assessment centre *OAR* validity, as the OAR is almost always used when assessment centres are used for selection purposes (as opposed to developmental purposes). Therefore, this study provides a meta-analytic update to the old value obtained by Gaugler *et al*. (1987). The current investigation is also based on a larger sample size ($N = 5850$) than the sample size of 4180 used in the Gaugler *et al*. (1987) meta-analysis.

The mean population estimate of the correlation between assessment centre OARs and supervisory performance ratings in the current study was $\rho = .28$ (95% CI $= .24 \leq \rho \leq .32$). Our estimate is thus significantly lower than the value of $\rho = .36$ (95% CI $= .30 \leq \rho \leq .42$) reported by Gaugler *et al*. (1987), which lies outside the 95% CI fitted around our

Table 1. Summary of meta-analysis results

Mean validity estimates for meta-analytic dataset					
N	*K*	*r*	r_c	r_{cb} (ρ)	95% CI
5850	27	.17	.20	.28	$.24 \leq \rho \leq .32$

Distribution of meta-analytic artifacts used to correct observed correlations								
U	σ^2_u	*c*	σ^2_c	r_{yy}	Σ^2_{ryy}	b	σ^2_b	$\bar{A}$
.94	.016	.94	.015	.52	0	.72	0	.64

Variance estimates for meta-analytic dataset				
σ^2_{rbc}	$\sigma^2_{e\_rbc}$	σ^2_ρ	Explained variance	SD_ρ
.0123	.0104	.019	84%	.04

Note: N, meta-analytic sample size; *K*, number of validity coefficients contributing to meta-analytic sample; *r*, mean sample weighted observed *r*; r_c, mean sample weighted observed *r* corrected for range restriction; r_{cb} (ρ), mean sample weighted observed *r* corrected for range restriction and criterion unreliability; *u*, mean range restriction on OARs; σ^2_u, variance in range restriction on OARs; *c*, mean range restriction correction factor; σ^2_c, variance in range restriction correction factor; r_{yy}, mean criterion reliability estimate; σ^2_{ryy}, variance in criterion reliability estimate; *b*, mean criterion unreliability correction factor; σ^2_b, variance in criterion unreliability correction factor; $\bar{A}$, mean artifact attenuation factor ($b \times c$); σ^2_{rbc}, variance in validities corrected for range restriction and criterion unreliability; $\sigma^2_{e\_rbc}$, weighted mean sampling error variance estimated for validities corrected for range restriction and criterion unreliability; σ^2_ρ, residual variance in validities corrected for range restriction and criterion unreliability; Explained variance, percentage of variance explained by sampling error; SD_ρ, standard deviation of corrected correlations.

estimated population value. A possible explanation for this finding is that the participants of modern assessment centres are subject to more pre-selection (given that they are so costly) than was customary in earlier assessment centres. This would result in more indirect range restriction in the modern assessment centres and consequently, in lower observed and corrected validities.

Unfortunately, we could not correct our data for indirect range restriction because the required indirect range restriction data were simply not reported in the primary studies. Nevertheless, in ancillary analyses we found some 'indirect' evidence of the impact of indirect range restriction on assessment centre data. Specifically, six studies within the meta-analytic dataset reported validities for cognitive ability tests that were used in the same selection stage within/alongside the assessment centre. The mean observed validity of these cognitive ability tests with respect to the criterion of job performance ratings was .10 ($N = 1757$). Thus, the validity of cognitive ability tests used *within or alongside an assessment centre* seemed to be much lower than the observed meta-analytic validities for cognitive ability tests as *stand alone predictors* (.24 and .22) reported by Hunter (1983) and Schmitt, Gooding, Noe, and Kirsch (1984) for US data. It is also much lower than the observed meta-analytic validity for cognitive ability tests as stand alone predictors (.29) on the basis of recent European data (Salgado *et al*., 2003). Although this comparison should be made with caution, it seems to indicate that the depressed validity of cognitive ability tests used *within/alongside* assessment centres might also result from considerable indirect range restriction on the predictor variable – most likely due to pre-selection on cognitive factors.

On a broader level, these results show that the selection stage should always be taken into account when reporting the validity of predictors (Hermelin & Robertson, 2001, see also Roth, Bobko, Switzer, & Dean, 2001). Hence, we urge future assessment centre researchers to routinely report (1) the selection stage within which assessment centres are used, (2) the pre-selection ratio of assessment centre participants, and (3) the correlation between the predictor composite used in preliminary selection stages and the OAR used in later stages. Only when this information becomes available, will it be possible to examine more fully the indirect range restriction issue in assessment centres and to perform the corrections for indirect range restriction [according to the procedures detailed in Hunter *et al*. (2006) and Schmidt *et al*. (2006)].

In regard to the potential presence of moderator variables, the current investigation suggests that very little variance remains unaccounted for once sampling error has been removed. This result contradicts the notion that assessment centres should show considerable variation in validities given the wide variations in their design and implementation. We believe that this result is more likely to be a result of little chance variation in the validity coefficients included in the dataset, rather than due to a genuine absence of moderator effects.

The following directions deserve attention in future research on the predictive validity of assessment centres. First, the criterion measures for validating assessment centres should be broadened. Over the last decade, one of the major developments in criterion theory is the distinction between task performance and citizenship performance (Borman & Motowidlo, 1993). To our knowledge, no studies have linked assessment centre ratings to citizenship behaviours. This is surprising because one of the key advantages of assessment centres is that they are able to measure interpersonally oriented dimensions. Second, it is of great importance that future studies examine the incremental validity of assessment centres over and above so-called low-fidelity simulations such as situational judgment tests (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). These tests have gained in popularity because of their ease of administration in large groups and low costs. In addition, they seem to capture interpersonal aspects of the criterion space and have shown good predictive validities.

Acknowledgement

We would like to thank Frederik Anseel for his insightful comments on a previous version of this manuscript.

References

Arthur, W., Jr, Day, E.A., McNelly, T.L. and Edens, P.S. (2003) A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions. *Personnel Psychology*, **56**, 125–154.

Borman, W.C. (1991) Job Behavior, Performance, and Effectiveness. In: Dunnette, M.D. and Hough, L.M. (eds), *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists Press, pp. 269–313.

Borman, W.C. and Motowidlo, S.J. (1993) Expanding the Criterion Domain to Include Elements of Contextual Performance. In: Schmitt, N. and Borman, W.C. and and Associates (eds), *Personnel Selection in Organizations*. San Francisco: Jossey-Bass, pp. 71–98.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. (1997) Bias in Meta-Analysis Detected by a Simple Graphical Test. *British Medical Journal*, **315**, 629–634.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. and Bentson, C. (1987) Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology*, **72**, 493–511.

Hermelin, E. and Robertson, I.T. (2001) A Critique and Standardization of Meta-Analytic Validity Coefficients in Personnel Selection. *Journal of Occupational and Organizational Psychology*, **74**, 253–277.

Hunter, J.E. (1983). *Test Validation for 12,000 Jobs: An application of job classification and validity generalization analysis to the general aptitude test battery*. USES Test Research Report No. 45, Division of Counseling and Test Development Employment and Training Administration, US Department of Labor, Washington, DC.

Hunter, J.E. and Schmidt, F.L. (1990) *Methods of Meta-Analysis: Correcting error and bias in research findings*. Beverly Hills, CA: Sage.

Hunter, J.E., Schmidt, F.L. and Le, H. (2006) Implications of Direct and Indirect Range Restriction for Meta-Analysis Methods and Findings. *Journal of Applied Psychology*, **91**, 594–612.

McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A. and Braverman, E.P. (2001) Use of Situational Judgment Tests to Predict Job Performance: A clarification of the literature. *Journal of Applied Psychology*, **86**, 730–740.

Roth, P.L., Bobko, P., Switzer, F.S. and Dean, M.A. (2001) Prior Selection Causes Biased Estimates of Standardized Ethnic Group Differences: Simulation and analysis. *Personnel Psychology*, **54**, 591–617.

Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. and Rolland, J.P. (2003) A Meta-Analytic Study of General Mental Ability Validity for Different Occupations in the European Community. *Journal of Applied Psychology*, **88**, 1068–1081.

Schmidt, F.L., Oh, I. and Le, H. (2006) Increasing the Accuracy of Corrections for Range Restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, **59**, 281–305.

Schmitt, N., Gooding, R.Z., Noe, R.A. and Kirsch, M. (1984) Meta-Analyses of Validity Studies Published Between 1964 and 1982 and the Investigation of Study Characteristics. *Personnel Psychology*, **37**, 407–422.

Thornton, G.C., III. (1992) *Assessment Centers and Human Resource Management*. Reading, MA: Addison-Wesley.

Viswesvaran, C., Ones, D.S. and Schmidt, F.L. (1996) Comparative Analysis of the Reliability of Job Performance Ratings. *Journal of Applied Psychology*, **81**, 557–574.

References to articles included in meta-analytic dataset

Anderson, L.R. and Thaker, J. (1985) Self-Monitoring and Sex as Related to Assessment Center Ratings and Job Performance. *Basic and Applied Social Psychology*, **6**, 345–361.

Arthur, W. and Tubre, T. (2001). *The Assessment Center Construct-Related Validity Paradox: An investigation of self-monitoring as a misspecified construct*. Unpublished Manuscript.

Binning, J.F., Adorno, A.J. and LeBreton, J.M. (1999). *Intraorganizational Criterion-Based Moderators of Assessment Center Validity*. Paper Presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, April.

Bobrow, W. and Leonards, J.S. (1997) Development and Validation of an Assessment Center During Organizational Change. *Journal of Social Behavior and Personality*, **12**, 217–236.

Burroughs, W.A. and White, L.L. (1996) Predicting Sales Performance. *Journal of Business and Psychology*, **11**, 73–84.

Chan, D. (1996) Criterion and Construct Validation of an Assessment Centre. *Journal of Occupational and Organizational Psychology*, **69**, 167–181.

Dayan, K., Kasten, R. and Fox, S. (2002) Entry-Level Police Candidate Assessment Center: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, **55**, 827–849.

Dobson, P. and Williams, A. (1989) The Validation of the Selection of Male British Army Officers. *Journal of Occupational Psychology*, **62**, 313–325.

Feltham, R. (1988) Assessment Centre Decision Making: Judgmental vs. mechanical. *Journal of Occupational Psychology*, **61**, 237–241.

Fleenor, J.W. (1996) Constructs and Developmental Assessment Centers: Further troubling empirical findings. *Journal of Business and Psychology*, **10**, 319–333.

Fox, S., Levonai-Hazak, M. and Hoffman, M. (1995) The Role of Biodata and Intelligence in the Predictive Validity of Assessment Centres. *International Journal of Selection and Assessment*, **3**, 20–28.

Goffin, R.D., Rothstein, M.G. and Johnston, N.G. (1996) Personality Testing and the Assessment Center: Incremental validity for managerial selection. *Journal of Applied Psychology*, **81**, 746–756.

Goldstein, H.W., Yusko, K.P., Braverman, E.P., Smith, D.B. and Chung, B. (1998) The Role of Cognitive Ability in the Subgroup Differences and Incremental Validity of Assessment Center Exercises. *Personnel Psychology*, **51**, 357–374.

Gomez, J.J. and Stephenson, R.S. (1987) Validity of an Assessment Center for the Selection of School-Level Administrators. *Educational Evaluation and Policy Analysis*, **9**, 1–7.

Higgs, M. (1996) The Value of Assessment Centres. *Selection and Development Review*, **12**, 2–6.

Hoffman, C.C. and Thornton, G.C., III. (1997) Examining Selection Utility Where Competing Predictors Differ in Adverse Impact. *Personnel Psychology*, **50**, 455–470.

Jones, R.G. and Whitmore, M.D. (1995) Evaluating Developmental Assessment Centers as Interventions. *Personnel Psychology*, **48**, 377–388.

McEvoy, G.M. and Beatty, R.W. (1989) Assessment Centres and Subordinate Appraisals of Managers: A seven year examination of predictive validity. *Personnel Psychology*, **42**, 37–52.

Moser, K., Schuler, H. and Funke, U. (1999) The Moderating Effect of Raters' Opportunities to Observe Ratees' Job Performance on the Validity of an Assessment Centre. *International Journal of Selection and Assessment*, **7**, 133–141.

Nowack, K.M. (1997) Congruence Between Self–Other Ratings and Assessment Center Performance. *Journal of Social Behavior and Personality*, **12**, 145–166.

Pynes, J. and Bernardin, H.J. (1992) Entry-Level Police Selection: The assessment center is an alternative. *Journal of Criminal Justice*, **20**, 41–55.

Robertson, I. (1999) *Predictive Validity of the General Fast Stream Selection Process*. Unpublished Validity Report, School of Management, UMIST.

Russell, C.J. and Domm, D.R. (1995) Two Field Tests of an Explanation of Assessment Centre Validity. *Journal of Occupational and Organizational Psychology*, **68**, 25–47.

Schmitt, N., Schneider, J.R. and Cohen, S.A. (1990) Factors Affecting Validity of a Regionally Administered Assessment Center. *Personnel Psychology*, **43**, 1–12.

Thomas, T., Sowinski, D., Laganke, J. and Goudy, K. (2005) *Is the Assessment Center Validity Paradox Illusory?* Paper Presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA, April.

Tziner, A., Meir, E.I., Dahan, M. and Birati, A. (1994) An Investigation of the Predictive Validity and Economic Utility of the Assessment Center for the High-Management Level. *Canadian Journal of Behavioral Science*, **26**, 228–245.

Appendix A

Table A1. Summary of validity results from studies included in meta-analysis

Author(s)	Observed overall validity coefficient	Sample size (*N*)	Range restriction in predictor (*U*)	Validity coefficient corrected for direct range restriction	Validity coefficient corrected for direct range restriction and criterion unreliability
Anderson and Thaker (1985)	.43	25	1	.43	.60
Arthur and Tubre (2001)	.23	70	.94	.24	.34
Binning, Adorno, and LeBreton (1999)	.15	1637	.8	.19	.26
Bobrow and Leonards (1997)	.23	71	1	.23	.32
Burroughs and White (1996)	.49	29	1	.49	.68
Chan (1996)	.06	46	1	.06	.08
Dayan, Kasten, and Fox (2002)	.17	420	.94	.18	.25
Dobson and Williams (1989)	.14	450	.55	.25	.35
Feltham (1988)	.16	128	.52	.30	.41
Fleenor (1996)	.24	85	1	.24	.33
Fox, Levonai-Hazak, and Hoffman (1995)	.33	91	1	.33	.46
Goffin, Rothstein, and Johnston (1996)	.30	68	.88	.34	.47
Goldstein, Yusko, Braverman, Smith, and Chung (1998)	.18	633	1	.18	.25
Gomez and Stephenson (1987)	.19	121	.94	.20	.28
Higgs (1996)	.32	123	.94	.34	.47
Hoffman and Thornton (1997)	.26	118	1	.26	.36
Jones and Whitmore (1995)	−.03	149	1	−.03	−.04
McEvoy and Beatty (1989)	.28	48	1	.28	.39
Moser, Schuler, and Funke (1999)	.37	144	1	.37	.51
Nowack (1997)	.25	144	1	.25	.35
Pynes and Bernardin (1992)	.23	68	1	.23	.32
Robertson (1999)	.23	105	.94	.24	.34
Russell and Domm (1995)	.22	140	1	.22	.31
Russell and Domm (1995)	.23	172	1	.23	.32
Schmitt, Schneider and Cohen (1990)	.08	402	.95	.08	.12
Thomas, Sowinski, Laganke, and Goudy (2005, April)	.30	56	.94	.31	.43
Tziner, Meir, Dahan, and Birati (1994)	.21	307	.94	.22	.31