Belbin Revisited: The Construct Validity of the Interplace II Team Role Instrument

Dirk van Dierendonck and Rob Groen

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2008-017-ORG
Publication	April 2008
Number of pages	29
Persistent paper URL	http://hdl.handle.net/1765/12123
Email address corresponding author	dvandierendonck@rsm.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website: www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES

RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	In the present study the construct validity of the revised edition of the Belbin Team Roles measure, the so-called Interplace II program, is tested. Three parallel parts were used to determine someone's team roles. The sample included 1434 persons who were asked to fill out the self-perception inventory and the self-perception assessment, whereas the observer assessment sheet was filled out by at least four observers. The inter-rater reliability appeared to be satisfactory across all team roles. As for the construct validity, which was studied in a multitrait-multimethod design using structural equation modeling, the results revealed that the discriminant and convergent validity for the instrument as a whole is good; only a small effect could be contributed to method variance.
Free Keywords	teams, Belbin, team roles, multitrait-multimethod methodology
Availability	
Classifications	

Belbin revisited: The construct validity of the Interplace II Team role instrument

Dirk van Dierendonck

RSM Erasmus University

Rotterdam

Rob Groen

CMB

Badhoevedorp

Work in Progress. Version april 1th 2008

Key words: Teams, Belbin, team roles, multitrait-multimethod methodology

RUNNING HEAD: Belbin revisited

The authors would like to thank Saake Buwalda and Ewald van den Hout for their help with the data collection. Correspondence may be sent to Dirk van Dierendonck, RSM Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Email: DvanDierendonck@rsm.nl

Abstract

In the present study the construct validity of the revised edition of the Belbin Team Roles measure, the so-called Interplace II program, is tested. Three parallel parts were used to determine someone's team roles. The sample included 1434 persons who were asked to fill out the self-perception inventory and the self-perception assessment, whereas the observer assessment sheet was filled out by at least four observers. The inter-rater reliability appeared to be satisfactory across all team roles. As for the construct validity, which was studied in a multitrait-multimethod design using structural equation modeling, the results revealed that the discriminant and convergent validity for the instrument as a whole is good; only a small effect could be contributed to method variance.

Belbin revisited: The construct validity of the Interplace II Team role instrument

With the increase of global competition there is a growing need for quick organizational processes. As a consequence, working in teams has more en more become a standard way of organizing (Kozlowski & Bell, 2004). Past research into the determinants of effective teams focused on various team issues, a central aspect being the influence of team structure, that is the number and, especially, the type of people that form the team. Diverse teams are believed to function better because as such they operate from a multiple-perspective view (Cox, Lobel, & McLeod 1991; Schippers, Den Hartog, Koopmam, & Wienk, J.A., 2003). The effects of different viewpoints in particular, caused by the wider range of information and expertise within teams, have been suggested to positively influence team performance (O'Reilly, Williams, & Barsade, 1998). This article addresses diversity by focusing on the different roles people may have within a team. In the past, several authors developed typologies of team roles (e.g. Davis, Millburn, Murphy, & Woodhouse, 1992, Spencer & Pruss, 1992). These so-called team roles are related to the personality characteristics individual members bring to the team. For example, there are indications that extraversion (Barrick & Stewart, 1997) and agreeableness (Neumann & Wright, 1999) are positively related to team effectiveness. Team member heterogeneity studies usually focus on characteristics as race, gender, career paths and educations (Stewart, 2006). Relatively little research, however, has been conducted into team role diversity, maybe as a result of the absence of an empirically validated instrument.

Probably, the most popular team role model within Europe is that developed by Belbin and his team (Belbin, 1981). The intuitive appeal and face validity made it quite popular among consultants and others professionally engaged in this field. However, it has been subject of academic criticism on its psychometric quality (Furnham, Steele, &

Pendleton, 1993; Broucek & Randell, 1996) ever since the so-called SPI, Belbin's most popular instrument, was introduced. A recent review of Belbin's team role model (Arizeta, Swailes, & Senior, 2007) provided mixed evidence on the convergent validity of the measure. Based on 43 empirical studies, these authors concluded that especially the discriminant validity of some of the scales is weak. To address this and other issues related to the validity and reliability of Belbin's measure, we set out to study the validity of an improved version of the original instrument. In doing so, we hope to show that this instrument warrants the attention of organizational psychologists interested in the effects of team diversity and team roles on team effectiveness. Furthermore, we hope it may give an impulse for future research in this particular field.

The Belbin team role model is the product of nine years of research, mostly conducted at the Administrative Staff College of Henley, by Belbin and his colleagues from the Industrial Training Unit from Cambridge (Belbin, 1981). Team effectiveness was studied in various different management games in which the composition of teams was manipulated in order to see how different personalities and abilities of team members contributed to team success. The model states that in addition to a professional and hierarchical role, team members also have a personality-bound team role. The participants in these games filled out Cattel's 16 PF questionnaire, Watson-Glaser Critical Thinking Appraisal and the Personal Preference Questionnaire. Originally, eight roles were distinguished, namely Implementer, Coordinator, Shaper, Plant, Resource-investigator, Monitor-evaluator, Team worker and Completer-finisher (see Table 1 for a short description of the roles). Later, a ninth role was added, namely that of Specialist. Belbin's theory states that out of these nine roles, every person has two or three team roles that fit naturally. The nine different roles are complementary. The effective team would ideally represent all nine team roles evenly

distributed over the team. There is some limited evidence supporting this proposition (Prichard & Stanton, 1999; Senior, 1997).

Belbin's model gained popularity after the publication of his book (Belbin, 1981) that included a Self-Perception Inventory of the original eight team roles, making it easy for people to determine their dominant team role. This self-perception inventory includes seven situations with eight possible reactions to each of them, each reaction corresponding to a specific team role. Ten points need to be distributed among the sentences in such a way that they give the best possible description of someone's behavior. The individuals are entirely free in the way they distribute the 10 points, all 10 points to just one sentence, all points evenly distribute, or anything in between. A few years later, the ninth team role was added to the SPI with the addition of a ninth possible reaction to each situation. Furthermore, a tenth possible reaction was added to each of the seven situations, a so-called dross item to measure social desirability. In addition, the Observer Assessment Sheet (OAS) was introduced. This is a peer-rater checklist with 72 adjectives that had to be filled out by people who know the individual. A software program (Interplace) was used to combine the answers on the SPI and the OAS and thus to calculate the team role scores. An individual would be given the scores on the nine team roles, together building a team role profile. Usually, their score would be higher on one or two team roles than on the other seven or eight, thus signifying their dominant team role.

In the late nineties of the last century a revised version of the instrument was developed in The Netherlands, the so-called the Interplace II team role instrument, in an attempt to deal with the criticism on Belbin's original instrument. In order to do so, The first step was the identification of persons who exemplified certain team role combinations. Similar to Belbin's original conceptualization, their scores were determined through their scores on general personality questionnaires, in this case the DCT, the 16 PF and the PPQ.

All individuals received feedback on their team role profile. Only those who fully identified with their profile remained in the study. Next, they participated in a training focusing on team roles with specific team role exercises. To remain in the final norm group, they had to behave according to their profile. This role model group consisted of 118 persons that were now being used to improve the Interplace instrument and to determine the formulas to calculate the team role scores. They had to fill out the SPI. Items that showed a team role inconsistent pattern were eliminated, new items were added, and some were partly modified. Three situations remained the same, two were changed, two were deleted, and three new ones were added. The subjects had to fill out the OAS for themselves. The Interplace II instrument that finally resulted from this consists of three parts: a revised SPI, an observation sheet with adjectives to be filled out by the person him- or herself (SPAS), and an observation sheet to be filled out by at least 4 colleagues. A computer program calculates team role scores for each part, and provides a final score weighing the self-reports for 50% and the observers scores for 50%.

Despite the existence of the Interplace program, with the exception of two studies (Broucek & Randell, 1996; Senour & Swailes, 1998), most research on the model has focused on the SPI only, hereby neglecting the observation sheets (e.g., Balderson & Broderick, 1996; Swailes & Aritzeta, 2006). Belbin himself noted that the early SPI is obsolete and should not be used by itself. The most important reason is that there is no way to counter illusions about the self without Observer Assessments. The latest edition of his book (Belbin, 2004) no longer contains the SPI. As this instrument is currently being used in consultancy and practice, research that both includes the self-report parts and the observation form is necessary in order to test the validity of this instrument.

The parallel parts to assess team roles provide an excellent opportunity to test the discriminant and convergent validity of this instrument. Regretfully, the first studies using

both parts in the original English version are not encouraging. Broucek and Randell (1996) tested the validity in a sample of 152 managers. Although the correlations between the same roles in the two measures displayed significant agreement, the average correlation was only .27. Also, correlations between the different roles of the same instrument were below expectation. In the same article, a sample of 123 persons filled out self-reports on both the SPI and the observer assessment sheet. Here the average correlation between the team roles was higher, that is .42. Considering that all scores were based on self-report data, one would expect higher correlations. Senior and Swailes (1998) tested the convergent validity in a sample of 65 individuals attending management courses. Only three out of nine correlations among the team role scores appeared to be significant. The correlations also raised some doubt about the discriminant validity because of the relative high correlations between certain team roles.

It should be noted that the team roles are not independent. For one thing, most people not only have a primary team role but also one or two secondary team roles. Moreover, according to Fisher, Hunter, & Macrossen (1998), the team roles can be subdivided into two categories: task oriented and relation oriented team roles. The task oriented team roles are: Monitor-evaluator, Plant, Shaper, and Completer-finisher. The relation oriented team roles are Resource-investigator, Coordinator, Team worker, and Implementer. A different subdivision was suggested by Belbin (1981): a 4 x 2 typology. Belbin combines two kinds of negotiators (Resource-investigator * Team worker), two kinds of intellectuals (Monitor & Plant), two kinds of managers (Implementer & Completer-finisher) and two kinds of team leaders (Coordinator & Shaper). The specialist was not part of either model. The studies into the underlying structure (Furnham, Steel, & Pendleton, 1993; Senior, 1998) confirmed neither of the models. Despite some similarity between the studies, the results were mostly

inconsistent. Therefore, more insight into the underlying structure of the Belbin team roles is still required.

In conclusion, the primary research focus of this article is the construct validity of the revised instrument. We will focus on the reliability of the observation sheet, and on the discriminant and convergent validity of the underlying team role dimensions as measured by the three parallel parts within the Interplace program.

Method

Subjects

The dataset consisted of 1434 individuals. All individuals filled out the self-perception list and had at least four other persons fill out the observation list. The sample included 972 (68%) men and 462 (32%) women representing. a cross section of professions from profit and non-profit organizations. It was a convenience sample of persons indicating interest in to receive their Belbin teamrole profile. The age of the persons was unknown.

The total number of observers was 6702. The number varied between 4 and 18 observers for each person. The observers could be colleagues, supervisors, or friends. The majority of the participants (1002) had four observers; 209 persons had 5 observers; 101 persons had six observers; 115 between 7 and 18 observers.

Measures

Team roles. The team roles were calculated with the Dutch version of the Interplace program: Interplace II Team Role Instrument. This version is based on the original English version developed by Belbin and colleagues (Belbin, 1993). The programme calculates team role scores based on the scores of the self-perception inventory, the self-perception adjectives and observation adjectives. Each person receives a score of 0, 1, 2 or 3 on each of the nine

team roles for each of the three parts separately. The final team role scores are based for 50% on the self-perception scores and for 50% on the observation scores.

Self-Perception Inventory (SPI) is an ipsative survey consisting of eight sentences describing a specific situation, followed by 10 choices of possible behavior in that situation. Nine of these choices are characteristic behavior of one of the team roles, the tenth choice is a social desirability item. Respondents are asked to divide 10 points among these choices. It is possible to divide these points over ten choices, give one choice a score of 9, or anything in between. The Interplace software calculates the team role indication based on these answers. When calculating the team role scores the social desirability scale is deleted. An important problem with a covariance matrix based on ipsative data is that it dooes not have an interpretable covariance matrix because of the constant-sum constraint (Cheung & Chan, 2002). By deleting the scores of the social desirability scale the SPI is no longer fully ipsative. A check of the resulting covariance matrix confirmed this because the sum of covariances of the team roles did not equal a constant, which is seen as the most problematic characteristic of ipsative scores (Clemans, 1996).

Observer Assessment Sheet (OAS) is a 81-adjective-item peer rater checklist divided into two parts, the first consisting of 57 unipolair positive adjectives, the second of 24 unipolair negative adjectives. Each observer gives a score of 1 to those adjectives that best characterize the person. A score of 2 is given when it is very appropriate. At least eight positive adjectives need to be checked.

Self-Perception Assessment Sheet (SPAS) is a 81-adjective-item checklist divided into two parts, similar to the OAS. The first part consists of 57 unipolair positive adjectives, the second of 24 unipolair negative adjectives. Subjects are asked to give 1 point to those adjectives that best characterize him- or herself and 2 points to adjectives that are vey appropriate. At least eight positive adjectives need to be checked.

Results

Inter-observer reliability

The inter-observer reliability was calculated with the Kendall's W coefficient of concordance. This is a nonparametric test to determine the agreement between observers. The Kendall's W is calculated for each person and may vary from 0 (no agreement) to 1 (fully agreement). The mean value for the group as a whole was .56 (SD = .18; median = .58; minimum .08; maximum = .94). Generally, a median value around .60 is considered to indicate moderate to strong agreement (Siegel & Castellan, 1988). The median value of .58 is statistically significant ($p < 01$) and quite acceptable if one takes into account that a person's behavior changes in the presence of different persons. There should, of course, be overlap (it is the same person), but there may also be unique variance (there are different perspectives). Interestingly, the correlation between the number of observers and the Kendall's W is -.19, indicating that there is a small, yet significant drop in agreement as the number of observers increases.

A different way to gain insight into the inter-observer agreement is the intra-class correlation. This correlation gives an indication of the proportion of variance at the second level (here the person). It can be interpreted as the expected correlation between randomly chosen observers of one person (Hox, 2002). Within SPSS, this intra-class correlation can be calculated by using the Mixed Model option and calculating the intercept only model. The correlations for each of the team roles are: Team worker = .33, Implementer = .42, Resource-investigator = .54, Monitor-evaluator = .38; Shaper = .47, Coordinator = .24, Completer-finisher = .35, Plant = .28, Specialist = .17. There is inter-rater agreement. The level of agreement, however, differs considerably depending on the team role. It seems that for certain behavioral patterns, most notably the extravert ones, like the Resource-investigator

and the Shaper, it is easier for observers to reach agreement than it is for the introvert roles like the Plant and certainly the Specialist.

Construct validity

The construct validity of the Interplace program is tested within a multitrait-multimethod (MTMM) design. Within such a framework it can be tested whether the three different methods (SPI, OAS, SPAS) concur in their assessment of the team role scores (convergent validity) and diverge in their measurement of the different team roles (discriminant validity) and to what extent method effects bias the results. Following Byrne (1998), we tested the MTMM design with covariance structure modeling, using Lisrel 8.71 (Joriskog & Sorbom, 2005). We also followed Byrne's (1998) guideline for testing the convergent and discriminant validity.

Four models were compared. The first model is the correlated traits/correlated methods model (see figure 1). It serves as a baseline to compare the other models and is composed of the nine team roles and the three method factors. All roles are allowed to correlate with each other. Similarly, the three method factors are allowed to correlate with each other. This model has a good fit (see table 3).

Subsequently, three models were tested. Model 2 is the no traits/correlated methods model. In this model, only method factors are specified. It has poor goodness-of-fit statistics. Model 3 is the perfectly correlated traits/freely correlated methods model. This model differs from model 1 in that the correlations between the traits are fixed at 1.0. In model 4, the freely correlated traits/uncorrelated methods model, the correlations between the method factors is specified at zero.

The convergent validity is tested by comparing the Chi-square goodness of fit indices of models 1 and 2: $\Delta\chi^2 = 9950{,}5$, $df = 64$, $p < .000$. This highly significant difference gives a

strong indication that independent ratings of the same team roles are correlated and supports the convergent validity of the three different measurement methods within Interplace. Further insight into the construct validity may be gained from examining the factor loadings and the factor correlations of Model 1 (see Tables 4 and 5). With respect to the convergent validity, the factor loadings of the team roles in Table 4 are all significant. The factor loadings of team roles are substantial, ranging from .50 to .82. The mean factor loading was highest for the OAS (.75), followed by the SPAS (.71), and lowest for the SPI (.60). The mean factor loadings on the method factors were lower compared to those on the team role factors, .32 for the SPI, .24 for the SPAS, and .21 for the OAS. This is another confirmation for the convergent validity of Interplace.

The discriminant validity is tested with respect to traits and methods. First, model 1 is compared with model 3 to see if independent measures of different roles are only negligible correlated. The larger the difference between the two models, the stronger the evidence for discriminant validity. The Chi-square goodness of fit indices were significantly different ($\Delta\chi^2$= 5178,49, df = 36, $p < .000$). This difference was quite large, that is the differences in the relative fit indices were substantial (ΔNNFI = .15, ΔCFI = .14) indicating strong discriminant validity. One may therefore conclude that, overall, with this instrument one can very well distinguish between the nine team roles. Nevertheless, the factor correlations between the team roles (Table 5) show that the discriminant validity of two team roles is not ideal. The Implementer role is highly correlated to Resource-investigator, Completer-finisher and the Plant (-.73, .71 and -.72, respectively). According to these results, people who score high as an Implementer, will also score high as a Completer/finisher, and low on Resource investigator and plant. The Resource investigator role is highly correlated with the Monitor/evaluator, the Shaper, and the Completer-finisher (.55, .67, -.73, respectively).

The possible influence of method variance can be determined by comparing model 1 and model 4. Here a relative small difference between the models speaks for a lack of influence of method variance on the results. Although the difference in Chi-square is significant ($\Delta\chi^2$= 84,86, df = 3, $p < .000$), the differences in relative fit indices are zero. This indicates that although method variance does play a role in the results, in practice its effect is only small.

With regard to the method used, the OAS gives the strongest indication of the team role. However, the method factor loadings also reveal method variance for a few roles in particular. The self-assessment methods give strong positive factor loadings for the Implementer (.55 and .49, respectively) and lowest negative factor loadings for the Plant (-.48 and -.51). This may indicate that with these self-assessment methods, subjects are biased in their answers, in that they prefer to consider themselves an Implementer rather than a Plant. The role of Monitor-evaluator is with .57 an outlier for the OAS method factor, indicating a greater tendency of being scored as a Monitor-evaluator by observers.

The correlations between the method factors show low correlations between the observation factor and the two self-perception factors. These different methods clearly provide different information on the team roles. More worrisome is the high correlation between the SPI and the SPAS, indicating a strong overlap in the information provided. It should be noted here that in calculating the end score on the nine team roles, the self-perception factors and the observation factor equally count for 50%. Our results can be interpreted as a confirmation of this practice.

An exploratory factor analysis with the combined team roles scores as produced by Interplace results in four factors with an eigenvalue greater than 1 (see Table 6). Factor 1 can be interpreted as the management factor with high factor scores of the Implementer and Completer finisher combined with negative factor scores of Resource-investigator and Plant.

The second factor as the critical follower with high positive scores on Monitor-evaluator and negative scores on Resource-investigator and Shaper. The third factor is the social factor combining high negative scores of team worker with moderate positive scores on Shaper and Specialist. The fourth factor is the generalist, combining high scores on the Coordinator with negative scores on the Specialist.

Discussion

The aim of this article was to study the construct validity of the revised version of Belbin's Team role program, Interplace II. Our results are encouraging. Most notably, we found a satisfactory inter-rater reliability across all team roles, for the instrument as a whole. Furthermore, we found good discriminant and convergent validity, whereas only a small effect could be attributed to method variance. Of course, there is still room for improvement for certain specific team roles. Although not unequivocally positive on all criteria and across all team roles, the results certainly show a more positive picture of the quality of this assessment instrument than previous studies would suggest.

First of all, the convergent validity across the three measurement methods is noteworthy. The factor loadings of model 1 show that all three methods individually contributed significantly to the nine team roles (see table 2). This is a strong and encouraging result given that Interplace uses two totally different self-report measures plus an observation measure that combines the ratings of at least four people. Since it was the reliability of the SPI that critics questioned (e.g. Furnham et al. 1993, Broucek & Randell, 1996), this is an important outcome in favor of the reliability of the measure. Given the partially ipsative nature of the SPI, it is doubtful how to interpret the results of these previous studies. To calculate the internal consistency, the researchers either had to 'create' data by setting al missing data on zero, or data had to be eliminated by only including those respondents that

divided their 10 points over all answers on a specific team role. In both cases the resulting values used to calculate the internal consistency are clearly different from the data the program itself uses to calculate team role scores. So, one can rightfully wonder what the real value of these previous studies is, whether the resulting values are unacceptably low (Furnham et al., 1993) or acceptable (Swailes & McIntyre-Bhatty, 2002, 2003). It was the partially ipsative nature of the SPI that led us to decide to work with the team role scores directly, thus avoiding the underlying measurement problems. It should also be noted that several authors showed that (partially) ipsative scores can be meaningfully factor analyzed (Saville & Willson, 1991; Ten Berge, 1999).

Secondly, the results showed support for the discriminant validity of the team role model as a whole. It clearly makes sense to differentiate among the nine team roles. Nevertheless, there are some high intercorrelations with the highest values between two latent variables of .73, indicating considerable overlap. In interpreting the correlation, please note that the correlations are between latent factors, the high correlations indicate similarity not that they are the same. One can speculate about this similarity. It may be the result of methodological indistinctness, or because certain team roles are by their very nature closely related. The second order factor analysis certainly points in this direction. Earlier theorizing also pointed towards the existence of underlying dimensions. The results of our exploratory second order factor analysis are very clear and add towards the validity of the measure. The management factor, the critical follower, the social factor, and the generalist are clearly behaviorally determined factors. Each signifies a different approach towards the issues people face in organizations in terms of observable behavior. These may be a more valid way to differentiate between the team roles than the ones suggested earlier which were more based on the focus of the behavior (task versus relationship) or the values underlying the behavior (Belbin's 4*2 typology).

Our results are more supportive than two earlier studies that compared the SPI with the OA (Broucek, & Randell, 1996; Senior & Swailes, 1998). These previous studies failed to find evidence for the convergent and discriminant validity of the two measures included. A possible explanation for this difference is that the revised version used in our study has enhanced the validity considerably.

Despite the weak points of the study, that is its cross-sectional nature and moreover the fact that we have as yet not solved the issue of the internal consistency of the scales, it has several strong points. First of all, the use of the MTMM methodology. Despite the fact that it has been introduced some time ago, it is seldom used by researchers, probably because of the difficulty to collect the necessary data. It is, however, an excellent methodology to study measurement issues, as was also recently shown by a study on measurement equivalence across rating sources (Woehr, Sheehan, & Bennett Jr., 2005). In this respect, the the three-way assessment of team roles is unique. Analyzed with the MTMM methodology, it gives an essential insight into the extent that team roles are recognized in a similar way between people. The fact that different self-report methodologies give much the same results gives confidence in the results. Furthermore, the large sample from a very diverse occupational background provides for possible generalization to other areas. It should also be acknowledged that in many studies, the SPI version used was the eight team role version in Belbin's original book, hereby neglecting the ninth team role (e.g. Arroba & Wedgwood-Oppenheim, 1994).

The most important practical implication is that our results stress the need to use the full instrument to determine team roles. Consultants should be aware that if they rely on the SPI alone, they run the real risk of an inaccurate insight into someone's dominant roles (to say the least). The strength of the Interplace program lies in the combination of different methodologies into one score, thereby controlling for the methodological weaknesses

inherent into each method. Even so, consultants using the program are well advised to be aware of the way this method may 'favor' certain team roles above others.

Now that we have an instrument that can give reliable and valid team role scores, more research is clearly needed. This methodology needs to be tested with other team role measures. Second, we need to know to what extent team roles are stable across time and circumstances. The inter-observer agreement of .56 suggests that their might be a stable and a variable element in team roles. Third, the one and only test for the Belbin Team role model still stands out, that is whether teams that have all team roles represented in their team indeed do perform better. This was Belbin's original premise, that, unfortunately, seldom has been tested (with possible encouraging exceptions of Prichard & Stanton, 1999; Rajendran, 2005). Some other studies on related topics (e.g. Aritza, Ayestaran, & Swailes, 2005; Fisher, Macrosson, Wong, 1998; Fisher, Macrosson, & Semple, 2000) have been conducted. The field certainly could do with more thorough investigations. However, with the early criticism on the self-report part of the instrument, extensive research has never been conducted, and as a result the underlying model has never really been tested. We hope that our results may encourage other researchers to include this team role measure in their studies.

References

Aritzeta, A., Ayestaran, S., and Swailes, S. (2005). Team role preferente and conflict management styles. *The International Journal of Conflict Management, 16*, 157-182.

Aritzeta, A., Swailes, S., and Senior, B. (2007). Belbin's team role model: Development, validity and applications for team building. *Journal of Management Studies, 44*, 96-118.

Arroba, T. and Wedgwood-Oppenheim, F. (1994). Do senior managers differ in the public and private sector? An examination of team role preferences. *Journal of Managerial Psychology,* 9, 13-16.

Balderson, S.J. and Broderick, A.J. (1996). Behavior in teams: Exploring occupational and gender differences. *Journal of Managerial Psychology, 11*, 33-42.

Barrick, M.R. and Stewart, G.L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology, 82*, 62-78.

Belbin, R.M. (1981, 2004). *Management teams: Why they succeed or fail*. London: Butterworth Heinemann Ltd.

Belbin, R.M. (1993). *Team Roles at work*. Oxford: Butterworth-Heinemann Ltd.

Broucek, W.G. and Randell, G. (1996). An assessment of the construct validity of the Belbin self-perception inventory and observer's assessment from the perspective of the five-factor model. *Journal of Occupational and Organizational Psychology, 69*, 389-405.

Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS*. Hillsdale, NJ: Lawrence Erlbaum.

Cheung, M.L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling, 9*, 55-77.

Clemans, W.V. (1996). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs, 14*, 1-56.

Cox, T.H., Lobel, S.A., & McLeod, P.L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal, 34*, 827-847.

Davis, J., Millburn, P., Woodhouse, M. (1992). *Successful team building: How to create teams that really work*. London: Kogan Page.

Fisher, M.G., Hunter, T.A., and Macrosson, W.D.K. (1998). The structure of Belbin's team roles. *Journal of Occupational and Organizational Psychology, 71*, 283-288.

Fisher, S.G., Macrosson, W.D.K., and Semple, J.H. (2000). Control and Belbin's team roles. *Personnel Review, 30*, 578-588.

Fisher, S.G., Macrosson, W.D.K., and Wong, J. (1998). Cognitive style and team role preference. *Journal of Managerial Psychology, 13*, 544-557.

Furnham, A., Steele, H., & Pendleton, D. (1993). A psychometric assessment of the Belbin team-role self-perception inventory. *Journal of Occupational and Organizational Psychology, 66*, 245-257.

Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahway, NJ: Lawrence Erlbaum Associates, Publishers.

Koslowski, S.W.J. and Bell, B.S. (2004). Work groups and teams in organizations. In: W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds) (2003). *Handbook of psychology: Industrial and organizational psychology, Vol. 12.* (pp. 333-375). Hoboken, NJ, US: John Wiley & Sons, Inc.

Neumann, G.A. and Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology, 84*, 376-389.

O'Reilly III, C.A., Williams, K.Y., and Barsade, S. (1998). Group demography and innovation: Does diversity help? In: D.H. Gruenfeld (Ed.). *Composition*. (pp. 183-207). US: Elsevier Science / JAI Press.

Prichard, J.S. and Stanton, N.A. (1999). Testing Belbin's team role theory of effective groups. *Journal of Management Development, 18*, 652-665.

Rajendran, M. (2005). Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin Self-perception Inventory. *Journal of Management Development, 24*, 738-753.

Savill, P. and Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology, 64*, 219-238.

Schippers, M.C., Den Hartog, D., Koopman, P.L., and Wienk, J.A. (2003). Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. *Journal of Organizational Behavior, 24***,** 779-802.

Senior, B. (1997). Team roles and team performance: Is there 'really' a link? *Journal of Occupational and Organizational Psychology, 70*, 241-258.

Senior, B. (1998). An empirically-based assessment of Belbin's team roles. *Human Resource Management Journal, 8*, 54-60.

Senior, B. and Swailes, S. (1998). A comparison of the Belbin Self Perception Inventory and Observer's Assessment Sheet as measures of an individual's team roles. *Internationl Journal of Selection and Assessment, 6***,** 1-8.

Siegel, S. and Castellan, N.J. jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Spencer, J. and Pruss, A. (1992). *Managing your team*. London: Piatkus.

Stewart, G.L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management, 32*, 29-54.

Swailes, S. and Aritzeta, A. (2006). Scale properties of the team role self-perception inventory. *International Journal of Selection and Assessment, 14,* 292-298.

Swailes, S. and McIntyre-Bhatty, T. (2002). The "Belbin" team role inventory: reinterprating reliability estimates. *Journal of Managerial Psychology, **17,*** 529-536.

Swailes, S. and McIntyre-Bhatty, T. (2003). Scale structure of the team role self perception inventory. *Journal of Occupational and Organizational Psychology, **76,*** 525-529.

Ten Berge, J.M.F. (1999). A legitimate case of component analysis of ipsative measures and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research, 34,* 89-102

Woehr, D.J., Sheehan, M.K., and Bennett Jr., W. (2005). Assessing measurement equivalence across rating sources: A multitrait-multirate approach. *Journal of Applied Psychology, 90*, 592-600.

Figure 1. Multitrait-multimethod confirmatory factor analytic model. The nine team roles, team worker (TW), implementer (IMP), resource investigator (RI), monitor evaluator (ME), shaper (SH), coordinator (co), completer finisher (CF), plant (PL), specialist (SP), are each rated by three different ratings sources, self-perception inventory (SPI), self-perception assessment sheet (SPAS), and observer assessment sheet (OAS).

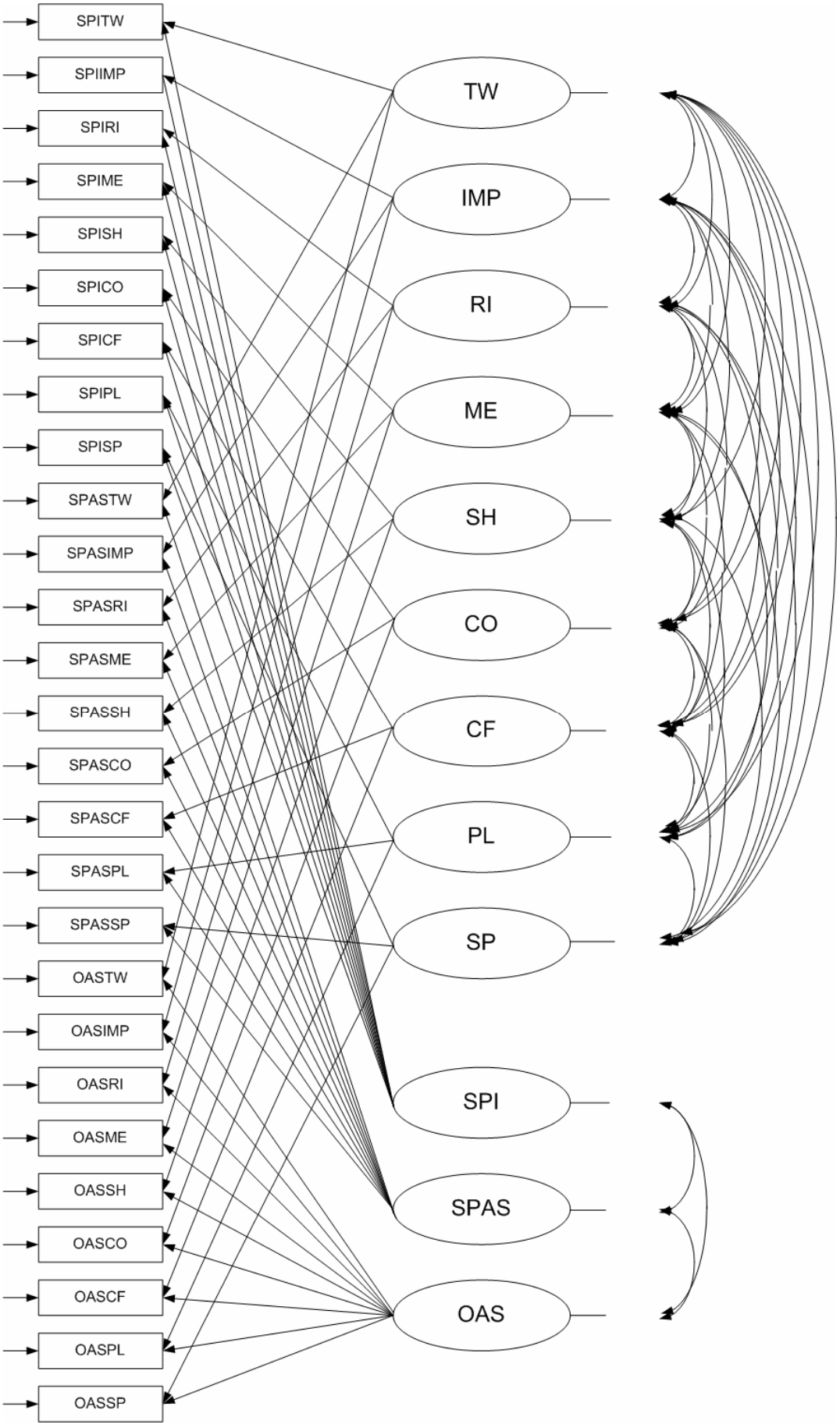
SPITW
SPIIMP
SPIRI
SPIME
SPISH
SPICO
SPICF
SPIPL
SPISP
SPASTW
SPASIMP
SPASRI
SPASME
SPASSH
SPASCO
SPASCF
SPASPL
SPASSP
OASTW
OASIMP
OASRI
OASME
OASSH
OASCO
OASCF
OASPL
OASSP
TW
IMP
RI
ME
SH
CO
CF
PL
SP
SPI
SPAS
OAS

Table 1

Belbin Team roles

Team role	Positive qualities	Allowable weaknesses
Team worker	An ability to respond to people and to situations, and to promote team spirit	Indecisiveness at moments of crisis
Implementer	Organizing ability, practical common sense, hard-working, self-discipline	Lack of flexibility. Unresponsiveness to unproven ideas
Resource Investigator	A capacity for contacting people and exploring anything new. An ability to respond to challenge	Liable to lose interest once to initial fascination has passed.
Monitor-evaluator	Judgement, discretion, hard-headedness	Lacks inspiration or the ability to motivate others
Shaper	Drive and a readiness to challenge inertia, ineffectiveness, compliancy or self-deception	Proneness to provocation, irritation, and impatience
Coordinator	A capacity for treating and welcoming all potential contributors on their merits and without prejudice	No more than ordinary in terms of intellect or creative ability
Completer-finisher	A capacity for follow-through. Perfectionism.	A tendency to worry about small things. A reluctance to ‘let go’
Plant	Genius, imagination, intellect, knowledge	Up in the clouds, inclined to disregard practical details or protocol
Specialist	Single-minded, self-starting, dedicated	Contributes on a narrow front only

Source: Belbin (1981, 1993).

Table 2. *Means, Standard Deviations, and intercorrelations for the Belbin Team Roles*

Team Roles	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1. SPITW	1.91	1.04	-																										
2. SPIIMP	1.51	1.17	.04	-																									
3. SPIRI	1.44	1.11	.08	-.56	-																								
4. SPIME	1.42	1.03	-.08	.00	-.05	-																							
5. SPISH	1.19	1.12	-.25	-.41	.53	-.20	-																						
6. SPICO	1.54	0.88	.18	.29	.06	.04	-.06	-																					
7. SPICF	1.19	1.07	.00	.56	-.57	-.05	-.28	.04	-																				
8. SPIPL	.77	.92	-.26	-.47	.36	.29	.34	-.28	-.33	-																			
9. SPISP	.60	.81	-.09	.30	-.39	.24	-.35	-.06	.29	.00	-																		
10. SPASTW	1.77	1.06	.54	-.02	.08	-.21	-.20	.08	-.04	-.21	-.06	-																	
11. SPASIMP	1.75	1.13	-.03	.61	-.48	-.00	-.34	.24	.49	-.46	.22	-.04	-																
12. SPASRI	1.41	1.13	.02	-.47	.62	-.25	.51	-.07	-.46	.28	-.34	.16	-.48	-															
13. SPASME	1.28	1.03	-.03	.08	-.22	.50	-.28	-.00	.09	.06	.20	-.18	.08	-.37	-														
14. SPASSH	1.06	1.07	-.32	-.27	.32	-.19	.56	-.12	-.21	.24	-.23	-.31	-.24	.51	-.32	-													
15. SPASCO	1.58	.99	-.06	.14	-.02	.09	.03	.38	-.02	-.18	-.14	-.06	.28	-.10	.14	-.02	-												
16. SPASCF	1.50	1.02	.28	.37	-.39	-.04	-.43	-.04	.45	-.31	.18	.32	.45	-.41	.18	-.40	-.14	-											
17. SPASPL	.91	1.06	-.09	-.47	.36	.16	.26	-.29	-.34	.56	-.11	-.12	-.58	.37	.17	.25	-.30	-.29	-										
18. SPASSP	.58	.71	-.22	.17	-.22	.07	-.08	-.11	.16	-.00	.35	-.15	.24	-.13	.08	.02	-.05	.07	.00	-									
19. OASTW	1.85	1.01	.47	-.00	.04	-.16	.23	.07	-.04	-.19	-.11	.53	-.04	.06	-.10	-.28	-.08	.24	-.07	-.18	-								
20. OASIMP	2.09	1.04	-.02	.47	-.41	.07	-.30	.16	.41	-.30	.22	-.06	.55	.45	.10	-.23	.16	.32	-.39	.13	-.08	-							
21. OASRI	1.39	1.14	.06	-.38	.50	-.27	.41	-.06	-.38	.18	-.35	.13	-.41	.62	-.34	.34	-.10	-.34	.26	-.18	.21	-.54	-						
22. OASME	1.17	1.04	-.02	.04	-.14	.52	-.28	.06	.00	.08	.19	-.11	.04	-.33	.54	-.30	.12	.07	.07	.08	-.11	.19	-.44	-					
23. OASSH	1.13	1.10	-.28	-.20	.25	-.22	.53	-.11	-.14	.18	-.23	-.25	-.19	.38	-.28	.61	-.01	-.34	.16	.01	-.33	-21	.48	-.40	-				
24. OASCO	1.42	.92	-.07	.10	.04	.07	.02	.33	-.-6	-.13	-.14	-.05	.17	-.03	.05	.02	.48	-.15	-.20	-.11	-.04	.29	-.07	.19	.01	-			
25. OASCF	1.52	1.02	.23	.33	-.40	.09	-.46	.03	.39	-.27	.20	.17	.36	-.45	.18	-.41	-.07	.55	-.24	.04	.30	.51	-.51	.28	-.49	-.06	-		
26. OASPL	.64	.88	-.08	-.40	.30	.16	.19	-.22	-.30	.45	-.08	-.08	-.43	.27	.10	.14	-.20	-.23	.56	.00	-.08	-.50	.30	.18	.15	-.26	-.27	-	
27. OASSP	.61	.74	-.19	.12	-.20	.12	-.07	-.17	.14	.06	.27	-.15	.12	-.14	.07	-.00	-.09	.03	.00	.41	-.20	.23	-.20	.11	.06	-.14	.08	.05	-

Note. SPI = self-perception inventory; SPAS = self-perception assessment sheet; OAS = observer assessment sheet; TW= team worker; IMP= implementer; RI=resource investigator; ME = monitor evaluator; SH = shaper; CO=coordinator; CF= completer finisher; PL= plant; SP= specialist

Table 3

MultiTraitMultiMethod models study 1 (n = 1432)

Model	X2	df	AIC	NNFI	CFI	SRMR
1. Correlated traits/ correlated methods	2354.86	258	2555.19	.92	.94	.06
2. No traits/correlated methods	12305.36	322	17105.05	.65	.68	.21
3. Perfectly correlated traits/ freely correlated methods	7533.35	294	11681.68	.77	.80	.11
4. Freely correlated traits/ Uncorrelated methods	2439.72	261	2577.73	.92	.94	.07

Table 4

Completely Standardized parameter estimates for factor loadings. MTMM, model 1.

		Method			Team roles							
		SPI	SPAS	OAS	TW	IMP	RI	ME	SH	CO	CF	PL	SP
SPI													
	TW	.07			.66								
	IMP	.55				.58							
	RI	-.47					.64						
	ME	-.21						.68					
	SH	-.31							.70				
	CO	.18								.53			
	CF	.41									.50		
	PL	-.48										.53	
	SP	.17											.57
SPAS													
	TW		.08		.73								
	IMP		.49			.69							
	RI		-.29				.78						
	ME		-.05					.77					
	SH		-.20						.76				
	CO		.18							.65			
	CF		.27								.70		
	PL		-.51									.68	
	SP		.11										.59
OAS													
	TW			-.07	.74								
	IMP			.17		.82							
	RI			-.26			.80						
	ME			.57				71					
	SH			-.20					.77				
	CO			.14						.71			
	CF			.25							.79		
	PL			.17								.82	
	SP			.07									.58

Note. SPI = self-perception inventory; SPAS = self-perception assessment sheet; OAS = observer assessment sheet; TW= team worker; IMP= implementer; RI=resource investigator; ME = monitor evaluator; SH = shaper; CO=coordinator; CF= completer finisher; PL= plant; SP= specialist

Table 5

Method and role correlations for MTMM, model 1

	Method			Team roles								
	SPI	SPAS	OAS	TW	IMP	RI	ME	SH	CO	CF	PL	SP
SPI	1.00											
SPAS	.69	1.00										
OAS	-.17	-.04	1.00									
TW				1.00								
IMP				-.11	1.00							
RI				.22	-.73	1.00						
ME				-.20	.17	-.55	1.00					
SH				-.49	-.38	.67	-.52	1.00				
CO				-.06	.39	-.05	.16	.01	1.00			
CF				.37	.71	-.73	.24	-.67	-.19	1.00		
PL				-.18	-.72	.44	.20	.29	-.47	-.44	1.00	
SP				-.37	.40	-.49	.32	-.17	-.31	.25	.03	1.00

Note. SPI = self-perception inventory; SPAS = self-perception assessment sheet; OAS = observer assessment sheet; TW= team worker; IMP= implementer; RI=resource investigator; ME = monitor evaluator; SH = shaper; CO=coordinator; CF= completer finisher; PL= plant; SP= specialist

Table 6

Factor loading exploratory factor analysis Team role scores

	I	II	III	IV
TW			-.90	
IMP	.87			
RI	-.66	-.54		
ME		.92		
SH		-.65	.44	
CO				.90
CF	.71			
PL	-.80			
SP			.54	-.45

Note. Only factor loadings > .40 are depicted. TW= team worker; IMP= implementer; RI=resource investigator; ME = monitor evaluator; SH = shaper; CO=coordinator; CF= completer finisher; PL= plant; SP= specialist

Publications in the ERIM Report Series Research* in Management

ERIM Research Program: "Organizing for Performance"

2008

Explaining Preferences and Actual Involvement in Self-Employment: New Insights into the Role of Gender
Ingrid Verheul, Roy Thurik and Isabel Grilo
ERS-2008-003-ORG
http://hdl.handle.net/1765/10979

Public Finance in China since the Late Qing Dynasty
Barbara Krug
ERS-2008-005-ORG
http://hdl.handle.net/1765/11287

Overoptimism among Founders: The Role of Information and Motivation
Ingrid Verheul and Martin Carree
ERS-2008-008-ORG
http://hdl.handle.net/1765/11557

Do Foreign Greenfields Outperform Foreign Acquisitions or Vice Versa? An Institutional Perspective
Arjen H.L. Slangen and Jean-François Hennart
ERS-2008-009-ORG
http://hdl.handle.net/1765/11558

Unemployment Benefits Crowd Out Nascent Entrepreneurial Activity
Philipp Koellinger and Maria Minniti
ERS-2008-012-ORG
http://hdl.handle.net/1765/11808

Acceleration of Technology Adoption within Firms: Empirical Evidence from the Diffusion of E-business Technologies
Philipp Koellinger and Christian Schade
ERS-2008-013-ORG
http://hdl.handle.net/1765/11809

Belbin Revisited: The Construct Validity of the Interplace II Team Role Instrument
Dirk van Dierendonck and Rob Groen
ERS-2008-017-ORG
http://hdl.handle.net/1765/12123

China's Institutional Architecture: A New Institutional Economics and Organization Theory Perspective on the Links between Local Governance and Local Enterprises
Barbara Krug and Hans Hendrischke
ERS-2008-018-ORG
http://hdl.handle.net/1765/12191

Ethics Programs and Ethical Cultures: A Next Step in Unraveling their Multi-Faceted Relationship
Muel Kaptein
ERS-2008-020-ORG
http://hdl.handle.net/1765/12192

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship