European Journal of Work and Organizational Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713684945

Belbin revisited: A multitrait-multimethod investigation of a team role instrument

Dirk van Dierendonck[a]; Rob Groen[b]
[a] RSM Erasmus University, Rotterdam, The Netherlands [b] CMB, Badhoevedorp, The Netherlands

First published on: 30 July 2010

To cite this Article van Dierendonck, Dirk and Groen, Rob(2010) 'Belbin revisited: A multitrait-multimethod investigation of a team role instrument', European Journal of Work and Organizational Psychology,, First published on: 30 July 2010 (iFirst)
To link to this Article: DOI: 10.1080/13594321003590580
URL: http://dx.doi.org/10.1080/13594321003590580

EUROPEAN JOURNAL OF WORK AND
ORGANIZATIONAL PSYCHOLOGY
0000, 00 (0), 000–000, i*First* article

Psychology Press
Taylor & Francis Group

Belbin revisited: A multitrait–multimethod investigation of a team role instrument

Dirk van Dierendonck
RSM Erasmus University, Rotterdam, The Netherlands

Rob Groen
CMB, Badhoevedorp, The Netherlands

In the present study, the construct validity of a revised edition of the Belbin Team Roles measure is tested. This edition consists of three parts 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 sheet, and the Observer Assessment Sheet was filled out by at least four observers. The interrater agreement of the Observer Assessment Sheet was satisfactory across all team roles. As for the construct validity, which was studied in a multitrait–multimethod design using structural equation modelling, the results revealed that the discriminant and convergent validity for the instrument as a whole is good; a small effect could be contributed to method variance.

Keywords: Teams; Methodology; Measurement development.

In modern organizations, working in teams has increasingly 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, which is the number and, especially, the type of people that form the team. The multiple-perspective view of members within diverse teams can lead to better and worse functioning teams (Cox, Lobel, & McLeod, 1991; Jackson, 1992; Schippers, Den Hartog, Koopman, & Wienk, 2003). One approach to studying the influence of team composition on team performance is by focusing on the different roles people may have within a team, their so-called team roles. In

Correspondence should be addressed to Dirk van Dierendonck, RSM Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: DvanDierendonck@rsm.nl

The authors would like to thank Saake Buwalda and Ewald van den Hout for their help with the data collection.

http://www.psypress.com/ejwop DOI: 10.1080/13594321003590580

the past, several authors developed typologies of team roles (e.g., Davis, Millburn, Murphy, & Woodhouse, 1992; Spencer & Pruss, 1992). However, team member heterogeneity studies usually focus on characteristics as race, gender, career paths, and education (Stewart, 2006). To encourage more research specifically into team roles, the availability of an empirically validated instrument would be very helpful. This article addresses the construct validity of an instrument to assess these team roles based on Belbin's team role model (Belbin, 1981). Specifically, this is the first study that tests the full Belbin team role instrument as it is currently used within consultancy practice within The Netherlands. It is also one of the first studies that focuses on the combined information provided by using both the self-perception and the observer part of the instrument to estimate a persons' team role.

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 due to problems with the psychometric quality of Belbin's most popular instrument, the Self-Perception Inventory (Broucek & Randell, 1996; Furnham, Steele, & Pendleton, 1993). A recent review of Belbin's team role model (Aritzeta, 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 a revised version of the original instrument using confirmatory factor analysis applied to multitrait–multimethod data.

THE BELBIN TEAM ROLE MODEL

The Belbin team role model is the product of 9 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 an interpersonal-oriented team role. The participants in these games filled out Cattell's 16 PF questionnaire, the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980), and the Personal Preference Questionnaire (Edwards, 1959). 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. According to Belbin, the effective team would ideally represent all nine team roles within a 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 (SPI) of the original eight team roles, making it easy to determine someone's 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 statements in such a way that they give the best possible description of someone's behaviour. The individuals are entirely free in the way they distribute the 10 points, all 10 points to just one statement, all points evenly distributed, 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

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 the 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).

possible reaction was added to each of the seven situations that was unrelated to a team role, a so-called dross item to measure social desirability. To enhance the accuracy of the instrument, the Observer Assessment Sheet (OAS) was introduced. This is a peer-rater checklist with 72 adjectives that is filled out by people who know the individual. A software program (Interplace) was used to combine the answers on the SPI and the OAS. 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.

Despite the existence of both the Self-Perception Inventory and the Observer Assessment Sheet, most research on the model has only focused on the SPI, hereby neglecting the Observer Sheet (e.g., Balderson & Broderick, 1996; Swailes & Aritzeta, 2006). Belbin himself noted that the early SPI is obsolete and should not be used by itself (Belbin, 2004). 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.

With the introduction of the independent methods used to assess team roles, we have 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 with respect to the convergent validity of the SPI and the OAS. Broucek and Randell (1996) tested the convergent validity in a sample of 152 managers. Although the correlations between the same roles in the two measures displayed significant relationships, the average correlation was only .27. The two roles with the strongest overlap were Shaper ($r = .40$) and Completer-finisher ($r = .38$). The correlation of Coordinator was lowest and not significant ($r = .11$). The correlations for the other team roles were .24 for Team Worker, .27 for Implementer, .30 for Resource-investigator, .29 for Monitor-evaluator, .29 for Plant, and .19 for Specialist. In the same article, a second 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. Similar low correlations were reported by Senior and Swailes (1998), who tested the convergent validity in a sample of 65 individuals attending management courses. Five out of nine correlations among the team role scores appeared to be significant (ranging between .36 and .50). The average correlation, however, was only .29. Neither study provided the necessary data for a full test of the discriminant validity in that they did not report the intercorrelations between team roles within the same measure.

INTERPLACE II, A REVISED INSTRUMENT

In the 1990s 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. The first step was the identification of individuals 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 Distorted Concepts Test (DCT), the 16 PF, and the Personal Preferences Questionnaire. 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 exercise focused 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 people who were used to improve the Interplace instrument. First, they filled out the original version of the Self-Perception Inventory. Two items that showed a team role inconsistent pattern were eliminated, three new situations were added, and for two situations some of the statements were modified. Next, the subjects had to fill out the Observer Assessment Sheet for themselves and they were asked to have at least four persons fill out this Observer Assessment Sheet for them. In a sense, this group exemplified ideal examples of the nine team roles. Their averaged scores on the measures within the Interplace instrument were therefore taken as most closely representing these team roles. The information from the role model group was now used to determine the formulas that are used to calculate the team role scores within the revised version of the software program.

The Interplace II instrument that resulted from this development process consists of three parts: a revised SPI, an observer sheet with adjectives to be filled out by the person him- or herself (SPAS), and an observer sheet to be completed by at least four colleagues. The four most important improvements of the Interplace II on the original Interplace instrument are: (1) The SPI has eight situations; only three situations are directly from the old SPI, the remaining five are adjusted versions. (2) The Observer Assessment Sheet is also filled out by the person him- or herself to improve the reliability of the self-perception estimate of one's team roles. (3) Nine adjectives were added to original 72 in the assessment sheet. (4) A differentiated scoring system was introduced whereby the answers were given weights –1, 0, 1, 2 on the team roles. A computer program uses the scores on the answers to calculate team role scores for each of the three parts, and provides a final score weighing the two self-report scores for 50% and the observers' scores for 50%. The 50/50 division is in a sense arbitrary. It was chosen to give information from both self-report and observer scores similar weight in the team role profile.

It is this instrument that is currently being used in consultancy and practice within The Netherlands. To test the validity of this instrument, research needs to include the self-report parts and the observer form. Insights gained from most earlier studies are—in a sense—obsolete because the version (i.e., the old SPI) they refer to is no longer in use by consultants working with the Interplace II program.

The psychometric quality and validity of the revised version of this team role instrument is explored in two steps. First, the interobserver agreement and reliability of the Observer Assessment Sheet is studied by analysing the extent that different observers agree in their assessment of the same person. Team roles are expected to be relatively stable across situations. Nevertheless, given that participants were free in their choice of observers, as long as these observers were aware of the interpersonal functioning of the participant, the interrater agreement is expected to be only moderately strong. Second, the convergent and discriminant validity is studied using the multitrait–multimethod (MTMM) design (Campbell & Fiske, 1959) using a covariance confirmatory factor analysis approach. With this approach, the variance explained by the methods can be differentiated from the variance explained by the team roles. With convergent validity, we study the extent that the three different methods within the Interplace II program (SPI, OAS, SPAS) concur in their assessment of the nine team roles. With discriminant validity, we study whether the nine team roles correlate with each other, at most, moderately. In addition, MTMM provides insight into the extent that method variance plays a role, that is the extent that correlations between team roles are determined by the measurement method instead of by conceptual overlap.

In conclusion, the primary research focus of this article is the validity of Belbin's team role model as measured by a revised version of the Interplace instrument. We will focus on the interobserver agreement and reliability of the observer sheet, and on the discriminant and convergent validity of the underlying team role dimensions as measured by the three methods within the Interplace program.

METHOD

Subjects

The dataset consisted of 1434 individuals. All individuals filled out the self-perception measures and had at least four other persons fill out the Observer Assessment Sheet. The sample included 972 (68%) men and 462 (32%) women representing a cross-section of professions from profit and nonprofit organizations. It was a convenience sample of persons indicating interest to receive their Belbin team role profile as part of a training focused on team

functioning. Therefore, only limited biographical information is available. The exact age of the participants was unknown; the majority of the sample was between 25 and 40 years old.

The total number of observers was 6702. The number varied between 4 and 18 observers for each person. The observers could be colleagues, their supervisor, or other persons who knew this person in the work context (e.g., clients). The majority of the participants (1002) had four observers; 209 persons had five observers; 101 persons had six observers; and 115 between seven and eighteen 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 program calculates team role scores based on the scores of the self-perception inventory, the self-perception adjectives of the self-perception assessment sheet and observer adjectives of the Observer Assessment Sheet.

Self-Perception Inventory (SPI). This is a survey consisting of eight sentences describing a specific situation, followed by 10 choices of possible behaviour in that situation. Nine of these choices are characteristic behaviour of one of the team roles; the tenth choice is a social desirability item. Examples of reactions for each of the team roles are for Team worker: ''I can get along with all kinds of people''; for Implementer: ''One can count on that I finish the tasks I take on''; for Resource-investigator: ''I quickly see new possibilities and take advantage of them''; for Monitor-evaluator: ''I make sure that decisions in the team are well thought-out''; for Shaper: ''I am willing to take a stand if that is essential to achieve important results''; for Coordinator: ''I quickly see when somebody can make a valuable contribution to the team''; for Completer-finisher: ''I make sure we do things that are realistic and attainable''; for Plant: ''I can come up with all kinds of original concepts to reach a solution''; for Specialist: ''I can contribute my specialized knowledge to a team''. 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 10, or anything in between.

An important problem with a covariance matrix based on ipsative data, like the SPI originally developed by Belbin, is that it does not have an interpretable covariance matrix because of the constant-sum constraint (Cheung & Chan, 2002). Within the revised version of Interplace that is tested in this article, this problem is resolved in different ways. To deal with

the possible negative effects of the ipsative nature of the SPI itself, first, before calculating the team role scores the social desirability scale is deleted. By deleting the scores of the social desirability scale the SPI is no longer fully ipsative. Second, the Interplace software uses the scores given to the remaining items to calculate the team role indication. There are two steps in this calculation. In the first step, for each item, a weighting factor based on the original norm group is determined for each team role. This factor can be zero (no relation to that particular team role), one, two, three, or minus one if a reaction is chosen that is a negative indicator for a team role. A weighting factor is used because the scores of the norm group showed that not all answers had the same relation to the team roles. Some answers could even be a contraindication of a team role. By using a weighting factor, a more accurate overall estimate could be calculated of each person's team role profile. In the second step, the scores on the separate answers were averaged for each of the nine team roles. Given that the weighting factors differ between the nine team roles, they can reach different minimum and maximum scores. For comparison purposes, this raw score is transformed into a score of 0, 1, 2, or 3. As a result, for all team roles, a score of zero indicated that a particular team role was not present and a score of three indicated the strong presence of a team role.

To sum up, for the SPI the inclusion of the social desirability item combined with the two steps used to calculate the team role scores makes sure that the resulting data that is used as input for the covariance matrix is no longer ipsative. The full interdependency that characterizes ipsative scores is solved by allowing for variance on the score of a team role that is not depending on that of the other team roles. A check of the resulting covariance matrix for the SPI team roles 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). This can be illustrated by checking an important characteristic of a covariance matrix based on ipsative data. When data is ipsative, the sum of the variance and covariances is zero for every row and column (Chan, 2003). Within our dataset, however, these sum scores were all above zero and differed between .35 and 1.03.

In addition, for the analysis it is important to realize that neither the OAS nor the SPAS are ipsative. The multitrait–multimethod analyses in this article are performed with the full covariance matrix consisting of covariances and variances of all three methods together.

Observer Assessment Sheet (OAS). This is an 81-adjective-item peer rater checklist divided into two parts, the first consisting of 57 unipolar positive adjectives, the second of 24 unipolar 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. Examples of adjectives for each of the team roles are for Team worker: tactful, helpful; for Implementer: precise, careful; for Resource-investigator: jovial, innovative; for Monitor-evaluator: thoughtful, reflective; for Shaper: pugnacious, daring; for Coordinator: decisive, consensus oriented; for Completer-finisher: well-organized, disciplined; for Plant: original, pondering; for Specialist: solo, specialized. The Interplace software calculates the team role indication based on these answers with a process similar to the SPI. At least eight positive adjectives need to be checked to get enough information for a reliable team role indication. The instructions asked for at least eight positive adjectives to make sure that the team role estimates would be based on the positive behaviour that a person showed, on the strengths that are part of each team role. Each person receives a score of 0, 1, 2, or 3 on each of the nine team roles.

Self-Perception Assessment Sheet (SPAS). This is an 81-adjective-item checklist divided into two parts, similar to the OAS. The adjectives are the same for the SPAS and the OAS. The first part consists of 57 unipolar positive adjectives, the second of 24 unipolar negative adjectives. Subjects are asked to give 1 point to those adjectives that best characterize him- or her and 2 points to adjectives that are very appropriate. At least eight positive adjectives need to be checked. The Interplace software calculates the team role indication based on these answers with a process similar as to the SPI. Each person receives a score of 0, 1, 2, or 3 on each of the nine team roles.

The final team role scores given as feedback are based on the team role scores for each of these three parts. The program provides a final score weighing the two self-report scores for 50% and the observers' scores for 50%.

RESULTS

Interobserver agreement and reliability

The first step in studying the psychometric quality of this team role instrument focuses on the interobserver agreement and reliability of the Observer Assessment Sheet. The interobserver agreement focuses on the extent that different observers give a similar indication of a person's team roles. This was calculated with the Kendall's W coefficient of concordance. This is a nonparametric test to determine the agreement between observers across all roles. The Kendall's W is calculated for each person and may vary from 0 (''no agreement'') to 1 (''full agreement''). The mean value for the group as a whole was 0.56 ($SD = 0.18$; median = 0.58; minimum 0.08;

maximum = 0.94). Generally, a median value around 0.60 is considered to indicate moderate to strong agreement (Siegel & Castellan, 1988). The median value of 0.58 is statistically significant ($p < .01$) and quite acceptable if one takes into account that a person's behaviour 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 of the same person). 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 way to gain insight into the reliability of the observer scores is the intraclass correlation (ICC1). 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). It gives an estimate of the relative consistency of the rank ordering of the observers. Within SPSS, this intraclass 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. Given these values, which are higher than the median of .12 reported by James (1982) for organizational literature, we can conclude that the reliability of the interrater scores is reasonable. The reliability, however, differs considerably depending on the team role. It seems that for certain behavioural patterns, most notably the extravert ones, like the Resource-investigator and the Shaper, observers are more likely to be consistent in their judgement than it is for the introvert roles like the Plant and certainly the Specialist.

Construct validity

The next step focuses on the construct validity of the three parts of the instrument by testing it 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 modelling, using Lisrel 2.72 (Jöreskog & Sörbom, 2005). The analysis was based on the scores on the nine team roles provided by the software. The intercorrelations for the team roles for all three methods can be found in Table 2. Given that these team role scores varied between 0 and 3 only, the continuity of the scores can be questioned and

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	0.77	0.92	*−.26*	*−.47*	*.36*	*.29*	*.34*	*−.28*	*−.33*	—																			
9. SPISP	0.60	0.81	*−.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	0.99	−.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	0.91	1.06	−.09	−.47	.36	.16	.26	−.29	−.34	**.56**	−.11	*−.12*	*−.58*	*.37*	*.17*	*.25*	*−.30*	*−.29*	—										
18. SPASSP	0.58	0.71	−.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	0.92	−.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	0.64	0.88	−.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	0.61	.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*	—

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. Monotrait–heteromethod correlations are in boldface; heterotrait–monomethod correlations are in italics.

they probably did not follow the assumption of a normal distribution. As such, a covariance matrix and an asymptotic weight matrix were computed from the ordinal responses with PRELIS 2.72 (Jöreskog & Sörbom, 2005) as input for the analysis with LISREL. An asymptotic matrix allows for weaker assumptions with regard to the distribution. Robust Maximum Likelihood was used as estimation method given its robustness to provide adequate goodness-of-fit statistics and standard errors (Bentler & Dudgeon, 1996). This estimation method has the added advantage that it provides the Satorra-Bentler scaled chi-square, which adjusts for the amount of kurtosis in the data to correct for the bias introduced when data are nonnormal in distribution. To calculate the difference between the models, the Satorra-Bentler scaled chi-square difference test is used (Satorra & Bentler, 2001).

We followed Byrne's (1998) guideline for testing the convergent and discriminant validity. Five 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) with comparative fit indices higher than .95 (NNFI = .95; CFI = .97). Also, the SRMR value is .08, which is usually seen as indicative of a good fit.

Subsequently, the no traits/correlated methods model is tested (Model 2). In this model, only method factors are specified. It tests to what extent overlap in variance between the team role scores can be explained by the specific measure used. This model has poor goodness-of-fit statistics, a first indication that method variance plays a limited role. The convergent validity is tested by comparing the corrected chi-square goodness of fit indices of Models 1 and 2, $\Delta\chi^2 = 8062.87$, $df = 64$, $p < .01$. 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 measures within Interplace. These three measures can also be divided into two groups according to source or according to instrument. For source, that is self-report (SPI and SPAS) versus observer (OAS) factors. For instrument, that is survey (SPI) versus assessment sheet (SPAS and OAS). To further our understanding of overlap in variance between the three measures, we analysed these two possible divisions as variations of Model 2. The fit of the source model was S-B$\chi^2 = 15542.72$, $df = 323$, $p < .01$, NNFI = .68, CFI = .70, SRMR = .16. The fit of the instrument model was S-B$\chi^2 = 15322.37$, $df = 323$, $p < .01$, NNFI = .68, CFI = .70, SRMR = .16. Although the fit increases somewhat from the original Model 2, neither model fits well to the data. This was not to be expected given the already poor fit of Model 2. There is a slightly better fit for the instrument model indicating that the specific method used has a slightly stronger

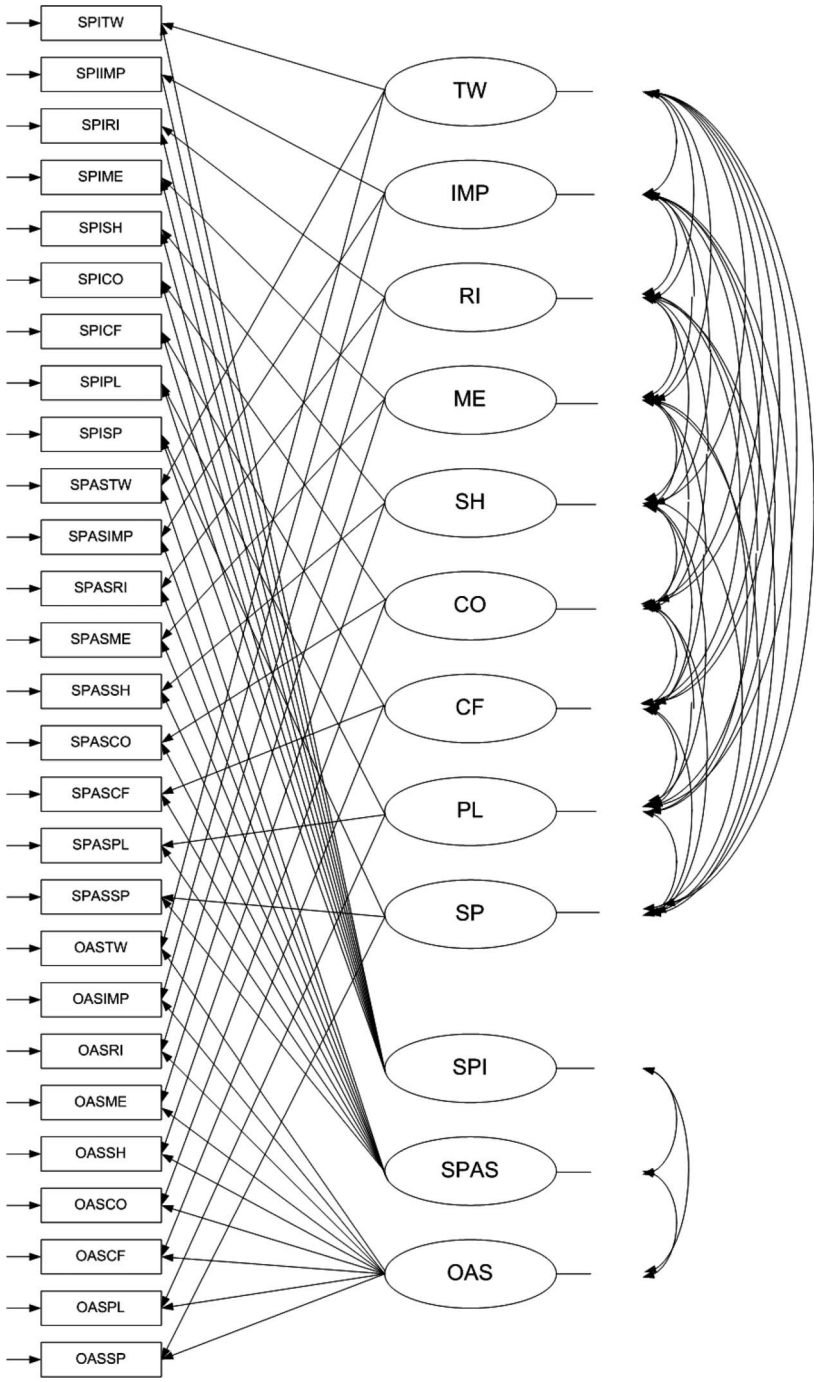

Figure 1. (Caption on next page).

TABLE 3
Multitrait–multimethod models, Study 1 ($n = 1434$)

Model	*S-Bχ2*	*df*	*AIC*	*NNFI*	*CFI*	*SRMR*
1. Correlated traits/correlated methods	2002.13	258	4724.78	.95	.97	.08
2. No traits/correlated methods	16626.32	322	25107.93	.65	.68	.26
3. Perfectly correlated traits/ freely correlated methods	9656.80	294	18220.38	.78	.82	.13
4. Correlated traits/no methods	3194.63	288	6881.07	.93	.94	.08
5. Freely correlated traits/ uncorrelated methods	2030.25	261	4724.56	.95	.97	.10

influence on overlap in scores than the source of the data. The influence of method variance on the results is more thoroughly discussed later.

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 the team roles are substantial on all three methods, ranging from .53 to .96. The mean absolute factor loading was highest for the OAS (.81), followed by the SPAS (.75), and lowest for the SPI (.64). 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 .26 for the OAS. The fact that the highest average factor loadings for the team roles was found for the observer scores is an indication of the value of including the OAS within Interplace to reach an accurate estimation of a persons' dominant team roles.

Next, the discriminant validity is explored by testing the perfectly correlated traits/freely correlated methods model (Model 3). This model differs from Model 1 in that the correlations between the traits are fixed at 1.0. The discriminant validity is tested with respect to traits and methods. To do so, Model 1 is compared with Model 3 to see if independent measures of different roles are only weakly correlated. The larger the difference between the two models, the stronger the evidence for discriminant validity. The corrected chi-square goodness of fit indices were significantly different, $\Delta\chi^2 = 47841.69$, $df = 36$, $p < .01$. This difference was quite large, the differences in the relative fit indices were also substantial (ΔNNFI = .17, ΔCFI = .15, ΔSRMR = .05) indicating strong discriminant validity. One may therefore conclude that, overall, with this instrument one can very well

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).

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	.12			.69								
IMP	.54				.65							
RI	−.41					.71						
ME	−.26						.69					
SH	−.30							.74				
CO	.25								.55			
CF	.33									.57		
PL	−.59										.53	
SP	.12											.66
SPAS												
TW		.09		.77								
IMP		.43			.77							
RI		−.26				.84						
ME		−.18					.83					
SH		−.20						.81				
CO		.24							.67			
CF		.20								.77		
PL		−.66									.69	
SP		.03										.64
OAS												
TW			−.09	.81								
IMP			.30		.86							
RI			−.36			.83						
ME			.55				77					
SH			−.26					.81				
CO			.20						.77			
CF			.35							.80		
PL			.08								.96	
SP			.14									.65

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.

distinguish between the nine team roles. Nevertheless, it should be acknowledged that 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 Plant (–.75, .72, and –.74, 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

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	.63	1.00										
OAS	−.14	.05	1.00									
TW				1.00								
IMP				−.12	1.00							
RI				.25	−.75	1.00						
ME				−.18	.20	−.57	1.00					
SH				−.48	−.37	.68	−.57	1.00				
CO				−.07	.38	−.04	.18	.02	1.00			
CF				.38	.72	−.73	.28	−.68	−.20	1.00		
PL				−.14	−.74	.46	.17	.28	−.44	−.46	1.00	
SP				−.37	.48	−.57	.33	−.24	−.31	.30	−.06	1.00

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.

Resource-investigator role is highly correlated with the Monitor-evaluator, the Shaper, the Completer-finisher, and the Specialist (.57, .68, –.73, and –.57, respectively).

The influence of method variance is studied with the correlated traits/no methods model (Model 4). The possible influence of method variance can be determined by comparing Models 1 and 4. The difference with Model 1 is that no method factors are specified in Model 4. The difference in chi-square is significant, $\Delta\chi^2 = 815.16$, $df = 30$, $p < .01$, indicating the influence of method variance. However, it should be noted that even in Model 4, the relative indices were already as high as .93, .94, and .08, respectively. This indicates that, although method variance does play a role in the results, in practice its effect might only be limited.

To understand the influence of method variance we turn first to Table 5, which also provides information on the correlations between the method factors. The correlations between both self-perception factors (SPI and SPAS) and the observer factor (OAS) is nonsignificant. The different perspective of self-rating versus observer rating clearly provides different information on the team roles. More worrisome, however, is the high, significant, correlation between the SPI and the SPAS, indicating a strong overlap in the information provided. Self-perception methods give strong overlap in the results that could be attributed to the use of similar source data.

Second, the method factor loadings depicted in Table 4 reveal the extent that team roles are over- or underestimated depending on the method used.

Table 4 provides the standardized parameters for the measurement model of Model 1 as depicted in Figure 1. This table provides insight into the extent that the variance of a team role score provided by one of three parts of Interplace is due to that specific method and due to the underlying latent team role factor. Positive parameters on one of the method factors indicate a tendency to overestimate; negative parameters indicate a tendency to underestimate. The self-assessment methods give strong positive factor loadings for the Implementer (.54 and .43, respectively) and strong negative factor loadings for the Plant (–.59 and –.66, respectively). 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. For the OAS method factor, the role of Monitor-evaluator is with .55 an outlier, indicating a greater tendency of being scored as a Monitor-evaluator by observers. These differences between self-perception and observer scores make sense given the different perspectives and insights of self compared to external observers. It should be noted here that in calculating the end score on the nine team roles, the self-perception factors and the observer factor equally count for 50%. Our results with respect to the influence of method variance can be interpreted as a confirmation of this practice.

DISCUSSION

The aim of this article was to study Belbin's team role model as measured by a revised instrument. Although we have to acknowledge that the results are 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—and therewith of the underlying model—than previous studies on Belbin's team roles have suggested. Most notably, for the Observer Assessment Sheet we found a satisfactory interrater agreement and reliability across all team roles. For the instrument as a whole, combining the three methods to assess team roles, 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. Overall, this is an important contribution to the academic literature on the Belbin team role instrument in that our results are more positive than most studies until now (Aritzeta et al., 2007).

Theoretical contributions

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 4). This is a strong and encouraging result given that Interplace uses two totally

different self-report measures plus an observer measure that combines the ratings of at least four people. This also is an important outcome in favour of the reliability of the measure. It dovetails encouraging results of a recent study by Swailes and Aritzeta (2006) who reported, using a large sample of 14,311 respondents, composite reliabilities for the scales of the original English version of the SPI between .63 and .78. It was the reliability of the Self-Perception Inventory that critics questioned (e.g., Broucek & Randell, 1996; Furnham et al., 1993). Given the partially ipsative nature of the SPI, it is doubtful how to interpret the results of previous studies. To calculate the internal consistency, the researchers either had to "create" data by setting all missing data to 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 (Swailes & McIntyre-Bhatty, 2002). 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).

Second, 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 these correlations, 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. Earlier theorizing also pointed towards the existence of underlying dimensions.

Our results are more supportive than two earlier studies that compared the Self-Perception Inventory with the Observer Assessment Sheet (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 so we do not know how stable team roles are over time 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 is a strength. Despite the fact that it has been introduced some time ago, it is seldom used by researchers, probably because of the difficulty of collecting 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, 2005). In this respect, the three-way assessment of team roles is unique. Analysed with the MTMM methodology, it provides 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 lends credence to the results. Another strong point is the large sample from a very diverse occupational background, which provides for possible generalization to other areas. It should also be acknowledged that, in many studies, the eight team role version in Belbin's original book was used, hereby neglecting the ninth team role (e.g., Arroba & Wedgwood-Oppenheim, 1994). Furthermore, we sidestepped the possible problems with the partially ipsative nature of the SPI by working with the transformed team role scores, thus avoiding the underlying measurement problems. By using the data provided by the Interplace program as input for our analyses, we stayed as close as possible to the way the instrument is actually used to determine a person's dominant team roles. In this way, the results provide insight into its practical validity. It should also be noted that several authors showed that (partially) ipsative scores can be meaningfully factor analysed (Saville & Willson, 1991; Ten Berge, 1999).

Practical implications

The most important practical implication is that our results emphasize 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 in each method. Even so, consultants using the program are well advised to be aware of the way this method may "favour" certain team roles above others. An open question is the number of observers needed for a valid and reliable estimate of a person's team roles. The results indicate a tradeoff between gathering enough different viewpoints for a complete estimate (i.e., in this case we choose for at least four observers), and the extent that observers have adequate knowledge of the behaviour of that person in a team. The small negative correlation between the number of observers and the Kendall's W indicates that it may not be necessary to overextend this number.

CONCLUSION

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 interobserver agreement of .56 suggests that there is 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, & Semple, 2000; Fisher, Macrosson, & Wong, 1998) 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., & Swailes, S. (2005). Team role preference and conflict management styles. *International Journal of Conflict Management*, *16*, 157–182.

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

Arroba, T., & 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., & Broderick, A. J. (1996). Behavior in teams: Exploring occupational and gender differences. *Journal of Managerial Psychology*, *11*, 33–42.

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

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

Belbin, R. M. (2004). *Management teams: Why they succeed or fail* (2nd ed.). London: Butterworth Heinemann.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, *47*, 563–592.

Broucek, W. G., & 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 Associates, Inc.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, *37*, 1–21.

Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, *30*, 99–121.

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., Murphy, T., & Woodhouse, M. (1992). *Successful team building: How to create teams that really work*. London: Kogan Page.

Edwards, A. L. (1959). *Edwards personal preference manual*. New York: Psychological Cooperation.

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

Fisher, S. G., Macrosson, W. D. K., & 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*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jackson, S. E. (1992). Team composition in organizational settings: Issues in managing an increasingly diverse workforce. I: S. Worchel, W. Wood, & J. A. Simpson (Eds.), *Group processes and productivity* (pp. 136–180). Newbury Park, CA: Sage.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, *67*, 219–229.

Jöreskog, K., & Sörbom, D. (2005). LISREL 2.72 [Computer software]. Chicago: Scientific Software.

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

Prichard, J. S., & 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.

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.

Savill, P., & 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., & 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., & Swailes, S. (1998). A comparison of the Belbin Self Perception Inventory and Observer's Assessment Sheet as measures of an individual's team roles. *International Journal of Selection and Assessment*, *6*, 1–8.

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

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

Swailes, S., & McIntyre-Bhatty, T. (2002). The ''Belbin'' team role inventory: Reinterpreting reliability estimates. *Journal of Managerial Psychology*, *17*, 529–536.

Swailes, S., & 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.

Watson, G., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal manual*. New York: Psychological Corporation.

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

Original manuscript received October 2008
Revised manuscript received December 2009
First published online month/year