

Acuerdo intragrupal: una aplicación a la evaluación de la cultura de los equipos de trabajo

José C. Sánchez y E. Alonso Amo*

Universidad de Salamanca y * Universidad Nacional de Educación a Distancia

El objetivo del presente trabajo es analizar cuál de los índices, el de James et al. (1984) o el de Lindell et al. (1999), ampliamente utilizados en la literatura, es el mejor para evaluar el acuerdo intra-grupal. Se realizaron dos estudios para tal fin. La variable que se consideró para evaluar el acuerdo intra-grupal fue la cultura de los equipos de trabajo. Aplicando la metodología y los criterios utilizados por Eby et al. (1999), ambos estudios demostraron que el índice de Lindell et al. es una estrategia más válida y apropiada para evaluar el acuerdo intra-grupal en las escalas con múltiples ítems que el de James et al.. Se discuten las implicaciones para la futura investigación y los métodos para estudiar los equipos.

Analysing the within-group agreement: An application to evaluate the culture of work teams. The aim of the present study is to analyse which of the indexes, the James et al.'s (1984) or the Lindell et al.'s (1999), widely used in the literature, is the best to evaluate the within-group agreement. Two studies were realized for such purpose. The variable that was thought to evaluate the within-group agreement was the culture of the work teams. Applying the methodology and the criteria used by Eby et al. (1999), both studies demonstrated that Lindell's index is a more valid and appropriate strategy to assess within-group agreement for multi-item scales than James et al.'s index. Implications for future research and methods for studying teams are discussed.

No resulta sorprendente afirmar que la investigación grupal no es tarea fácil. Tanto en el ámbito académico como en el profesional, los expertos han estado confusos con respecto a cómo evaluar las entradas (*inputs*), los procesos y los resultados (*outputs*) a nivel de análisis grupal. Los investigadores frecuentemente recogen datos a nivel de sujeto para operativizar luego constructos no sólo a nivel de individuo sino también a nivel organizacional y/o grupal. Por ejemplo, se evalúa primero y se agregan después las percepciones que los miembros de una organización tienen sobre el clima (análisis individual) para luego analizar el clima de la organización (análisis organizacional); o también los investigadores evalúan y suman las evaluaciones de los subordinados con respecto a su líder para operativizar el carisma de líder; etc. Algunos investigadores, no obstante, han comenzado a plantearse cómo medir mejor estos constructos que tienen –o pueden tener– un referente claramente grupal, como por ejemplo el constructo «eficacia colectiva» (cf. Gibson, Randel y Earley, 2000).

Los investigadores a la hora de recoger datos a nivel individual para operativizar constructos a nivel grupal u organizacional implícita o explícitamente proponen *modelos de composición* para sus constructos. Los modelos de composición «especifican la relación funcional entre fenómenos o constructos a diferentes niveles de análisis... que se refieren esencialmente al mismo contenido

pero que son cualitativamente diferentes a distintos niveles de análisis» (Chan, 1998, p.234). En la tipología explicitada por Chan sobre los modelos de composición, el acuerdo intra-grupal ocupa un lugar destacado, principalmente en los modelos de consenso directo, de consenso de cambio de referente y de dispersión. Es en los dos primeros donde el acuerdo intra-grupal es un prerequisite para la agregación de datos a nivel grupal. En estos modelos el acuerdo intragrupal es un elemento fundamental para poder definir y analizar un constructo a nivel grupal. El investigador asume que los miembros integrantes del grupo son homogéneos con respecto al constructo de interés. En el modelo de dispersión, por el contrario, los investigadores sugieren que los grupos varían en su nivel de homogeneidad. No obstante, tanto los modelos de dispersión como de consenso atribuyen el mismo significado al acuerdo intragrupal: el grado en que los miembros del grupo constituyen un grupo homogéneo, respecto a sus percepciones, intereses, conductas y otras características.

Para analizar los datos que se recogen en la evaluación de estos constructos grupales contamos con diferentes métodos; el más utilizado se basa en evaluar las percepciones de los miembros del grupo a cerca de características personales. A partir de aquí los procedimientos que se siguen varían de unos a otros. Un primer procedimiento consiste en pedir a los miembros del grupo que respondan a los ítems de un cuestionario y estas respuestas son luego promediadas para obtener un valor a nivel de grupo. Por ejemplo, las percepciones de auto-eficacia individual realizadas por cada miembro del grupo son agregadas para formar una puntuación grupal, o una medida de eficacia colectiva. No obstante, este enfoque ha sido criticado porque, por ejemplo, la agregación de las puntuaciones individuales a los ítems puede que no recoja, en algunos

Fecha recepción: 20-2-03 • Fecha aceptación: 17-6-03

Correspondencia: José C. Sánchez

Facultad de Psicología

Universidad de Salamanca

37005 Salamanca (Spain)

E-mail: jsanchez@usal.es

casos, el sentido colectivo de la habilidad del equipo para resolver con éxito las tareas de grupo (p.ej., Bandura, 1997; Gist, 1987; Lindsley, Brass y Thomas, 1995).

Un segundo procedimiento es similar al primero pero con la diferencia de que los ítems están redactados para recoger atributos del equipo como un todo. Así, el referente utilizado en estos ítems cambia del individuo al grupo. Continuando con el ejemplo de la eficacia, cada miembro del equipo proporcionaría su propia percepción de la capacidad del equipo (no de la suya propia), y estas puntuaciones son luego agregadas para formar una puntuación grupal (e.g., Earley, 1999; Feltz y Lirgg, 1998). Aunque existe evidencia de que este segundo procedimiento mejora el acuerdo intragrupal y fomenta la variabilidad intergrupala (Klein, Conn, Smith y Sorra, 2001), también ha sido criticado por no recoger fielmente los procesos a nivel de grupo porque el nivel de medición aún reside en el individuo y no en el grupo (Fleener, Fleener y Grossnickle, 1996).

En suma, tanto el primero como el segundo procedimiento se basan en la agregación. Y la principal crítica que ha recibido esta forma reside en que no permite la interacción entre los integrantes del grupo para responder a los ítems, lo que limita bastante el estudio de los procesos grupales subyacentes (Gibson et al., 2000).

Frente al método de agregación se ha propuesto como alternativa el método del consenso –o discusión grupal– (véase p.ej., Earley, 1999; Kirkman, Tesluk y Rosen, 2001). A diferencia de aquél, en los dos procedimientos antes comentados, este método supone que el grupo entero responde de forma colectiva mediante el consenso a los ítems diseñados para evaluar un constructo grupal. Así, el equipo proporciona una única respuesta para cada ítem y no hay necesidad de agregar las respuestas individuales. Desde la perspectiva del investigador esta metodología es bastante laboriosa. Por otro lado, hay poca evidencia hasta la fecha que demuestre la superioridad del método del consenso con respecto al de agregación. Kirkman et al. (2001) encontraron que el método del consenso para medir la delegación del poder (*empowerment*) del equipo explicó significativamente mayor varianza en la efectividad del equipo (medida con las variables productividad, servicio al cliente y proactividad) que el método de agregación. Hemos de considerar, no obstante, que el método del consenso es vulnerable a la pérdida de procesos grupales así como a todas las limitaciones reconocidas sobre toma de decisiones grupales, tales como pensamiento grupal, diferencia de estatus, conformidad, dominación de la discusión por uno o dos miembros, o el tiempo necesario para llegar al consenso.

El presente trabajo más que comparar el poder predictivo de cada uno de estos métodos se centra en comparar dos índices de acuerdo interjueces –el de James, Demaree y Wolf (1984, 1993) y el de Lindell, Brandt y Whitney (1999)– para ver cuál de ellos es el más adecuado para medir constructos que tienen claramente un referente grupal.

La medida que se ha venido utilizando para dar cuenta de este acuerdo intragrupal ha sido el índice de acuerdo interjueces de James, Demaree y Wolf (1984; 1993). Este índice (r_{wg}) compara el acuerdo promedio observado a nivel de ítem entre los individuos de un grupo con la varianza esperada de la población. James et al. (1984) nos proporcionan dos fórmulas, una para un único ítem [$r_{wg(i)}$] y otra para múltiples ítems [$r_{wg(j)}$]. La última fórmula es de interés para la presente investigación. El índice r_{wg} se ha utilizado principalmente para justificar la agregación de respuestas individuales a niveles de análisis superiores.

El índice propuesto por James et al. (1984) para múltiples ítems es:

$$r_{wg(j)} = \frac{J[1 - (\bar{S}_{xj}^2 / \sigma_{EU}^2)]}{J[1 - (\bar{S}_{xj}^2 / \sigma_{EU}^2)] + (\bar{S}_{xj}^2 / \sigma_{EU}^2)} \quad (1)$$

En donde J es el número de ítems de la escala; \bar{S}_{xj}^2 es la media de las varianzas observadas en los J ítems; σ_{EU}^2 es la varianza esperada del error (E) basada en una distribución uniforme (U), el término $\sigma_{EU}^2 = (A^2 - 1)/12$, en donde A es el número de alternativas posibles en una escala tipo Likert.

James et al. (1984) señalan que este índice se puede utilizar como un indicador del acuerdo perceptivo entre los individuos de un mismo grupo. Aunque la utilización del índice r_{wg} con este propósito, tiene sentido conceptualmente, el índice para múltiples ítems ($r_{wg(j)}$) puede mostrar un comportamiento irregular y puede no estar distribuido normalmente (Lindell y Brandt, 1997). Esto se debe a la corrección Spearman-Brown, que se pone en el índice $r_{wg(j)}$ de James et al., y se sugiere que una función lineal más simple podría utilizarse para calcular el grado de acuerdo intragrupal y evitar esta y otras propiedades no deseables (Lindell, 2001).

Lindell, Brandt y Whitney (1999) nos han proporcionado un índice revisado (r_{wg}^*) que no presenta una conducta irregular y que tiene la ventaja añadida de la linealidad. El índice revisado sencillamente quita la corrección Spearman-Brown de tal manera que la varianza media obtenida se compara con la varianza esperada bajo una distribución uniforme de las respuestas:

$$r_{wg(j)}^* = 1 - (\bar{S}_n^2 / S_{EU}^2) \quad (2)$$

Donde \bar{S}_n^2 es la varianza media a nivel de ítem promediada para todos los ítems de un grupo dado y S_{EU}^2 es la varianza esperada basada en una distribución derivada teóricamente (e.g., rectangular, uniforme, ligeramente asimétrica).

A pesar de la utilización indistinta de estos dos índices, la literatura está falta de los correspondientes estudios comparativos, si exceptuamos el trabajo de Eby, Meade, Parisi y Douthitt (1999). No obstante, estos autores, para establecer la superioridad de estos dos índices, se sirvieron de grupos formados por la aleatorización de una muestra de individuos, lo que a nuestro entender deja mucho que desear a la hora de establecer conclusiones válidas para los equipos de trabajo. El presente trabajo tiene como principal objetivo, por tanto, cubrir esta laguna. Hemos utilizado para ello una muestra de equipos reales de trabajo de diferentes organizaciones, lo que aporta una mayor validez ecológica en el análisis de estos índices. En el caso concreto que nos ocupa, hemos considerado la cultura de los equipos de trabajo como constructo grupal para poder comparar estos dos índices de acuerdo.

Primer estudio

Método

Muestra

Un total de 198 sujetos pertenecientes a 39 equipos de trabajo diferentes participaron en el estudio. El promedio de edad de los participantes fue de 35 años (sd= 8,27). El 51% fueron hombres y

el 49% mujeres. El 46,70% de los sujetos son empleados de empresas públicas y el 53,3% de empresas privadas. El 52% de los sujetos tenía título universitario. La antigüedad media en el equipo fue de 5,5 años (sd= 6,1) y en la empresa de 9,4 años (sd= 9,07).

Instrumento

Los datos se recogieron utilizando el cuestionario de cultura organizacional (C.N.O.-2). Este cuestionario representa la versión modificada del cuestionario CNO (Alonso, 1997; Sánchez y Alonso, 1997, 1998). El C.N.O.-2 consta de 82 ítems que intentan evaluar a partir de la información facilitada por la persona focal los siguientes aspectos de la cultura organizacional (subescalas): conductas de consideración; de innovación; de orientación al cambio; de orientación a la acción; calidad de trabajo; orientación a los resultados; colaboración; orientación hacia la autoridad; realización del rol; competición; tolerancia a la incertidumbre; integración de los grupos; control de la acción; y, libertad de acción. A los sujetos se les pedía que señalasen en una escala tipo Likert, en un rango de 0 de 4 puntos, siendo el 0 indicador de que la conducta es «nada característica» y el 4 «muy característica», «en qué medida esa conducta es característica de su lugar de trabajo».

Procedimiento

Una vez recogidas las puntuaciones de los sujetos a través del cuestionario de cultura y tras el análisis de fiabilidad del mismo procederemos a desarrollar las pruebas que proponen Eby, Meade, Parisi y Douthitt, (1999) para poder establecer la superioridad de estos dos índices. Estos autores consideran como propiedades deseables de una medida la superioridad en las siguientes pruebas psicométricas:

- a) *La variabilidad de las medidas:* Para ello calcularemos los acuerdos intragrupal utilizando las respectivas fórmulas para poder compararlos. Es importante indicar que el estadístico r_{wg} se ordena en un rango teórico de 0 a 1 y se puede interpretar como un coeficiente de fiabilidad (i.e., los valores de 0,70 o superiores se consideran un acuerdo de moderado a alto). En contraste, los valores r_{wg}^* se ordenan en un rango de +1 (máximo acuerdo) a -1 (máximo desacuerdo). Pero ya que el punto cero de los índices tiene un significado comparable, la regla a ojo del 0,70 se aplica a ambos índices.
- b) *La correlación entre la varianza media de los ítems y su correspondiente nivel de acuerdo intragrupal.* Puesto que el índice de la varianza media representa la dispersión media de las puntuaciones a través de todos los ítems de los miembros del equipo, es de esperar una relación inversa entre esas dos variables.
- c) *La normalidad en la distribución de las puntuaciones de los índices de acuerdo intragrupal.*
- d) *Los valores de la kurtosis*

Resultados

De los 82 ítems que integraban el total del cuestionario se excluyeron ocho ítems por no cumplir alguno de los criterios propuestos por Strealau, Angleitner, Bantelmann y Ruch (1990): a) el coeficiente de correlación de Pearson entre la puntuación en el

ítem correspondiente y la suma de las puntuaciones en los ítems restantes es menor a ,20; b) la correlación del ítem con su subescala es menor que la correlación con las otras subescalas; y c) la correlación del ítem con su subescala es significativa. A estos criterios nosotros consideramos que las subescalas deberían tener más de cinco ítems y un coeficiente de fiabilidad (alfa de Crombach) por encima de ,70 (Nunnally, 1978). Con estos dos últimos criterios descartamos las subescalas «libertad de acción» y «tolerancia a la incertidumbre». El resto de las subescalas presentaron unos valores de fiabilidad entre ,70 y ,88.

En la tabla 1 se presentan los valores obtenidos utilizando los dos índices de acuerdo, el de James et al. (1984) y el de Lindell et al. (1999). Como podemos observar, casi todos los valores de Rwg se agrupan en los rangos más altos, mientras que los valores de Rwg^* muestran considerablemente una mayor variabilidad.

La evidencia de la superioridad del índice Rwg^* también se puede comprobar examinando la correlación entre la varianza media de cada equipo a lo largo de todos los ítems y su correspondiente nivel de acuerdo. La correlación entre el valor Rwg^* para cada equipo y su correspondiente varianza promedio fue de $r = -1$ ($p < ,05$), tanto para la distribución uniforme como para la distribución ligeramente sesgada. Por el contrario, la correlación entre los valores Rwg para cada grupo y su correspondiente varianza promedio fue de $r = -,034$ y $r = ,202$, para la distribución uniforme y sesgada, respectivamente. Por tanto, los valores Rwg^* se corresponden más correctamente a las diferencias en la varianza media de las respuestas dentro del grupo, mientras que los valores Rwg sólo comparten el 0,01% y el 0,04% de la varianza en la distribución uniforme y sesgada, respectivamente.

Otra propiedad deseable de una medida es su normalidad. En este caso utilizamos el estadístico Kolmogorov-Smirnov, que compara las frecuencias empíricas o reales con las frecuencias teóricas o esperadas, si la distribución fuese normal. En la tabla 2 los índices Rwg^* , utilizando la distribución uniforme y sesgada, se aproximan a la normalidad, mientras que el índice Rwg sólo se

Tabla 1
Variabilidad de los índices de James et al (1984) y Lindell et al (1999)

Valores	Rwg^*		Rwg	
	Uniforme	Asimétrica	Uniforme	Asimétrica
de 1,01 al más alto			10	2,1
de ,9 a 1	4	,9	197	42,2
de ,8 a ,89	19	4,1	151	32,3
de ,7 a ,79	51	11,1	44	9,4
de ,6 a ,69	93	20,2	15	3,2
de ,5 a ,59	91	19,7	14	3,0
de ,4 a ,49	81	17,6	12	2,6
de ,3 a ,39	36	7,8	5	1,1
de ,2 a ,29	29	6,3	4	,9
de ,1 a ,19	27	5,9	4	,9
de ,0 a ,09	18	3,9	3	,6
De -,1 a -,19	4	,9	1	,2
De -,2 a -,29	3	,7	2	,4
De -,3 a -,39	4	,9	1	,2
de -,4 a -,49	1	,2	0	0
de -,5 a -,59			0	0
de -,6 a -,69			0	0
de -,7 a -,79			0	0
de -,8 a -,89			1	,2
de -,9 a -,99			5	1,1
de la más baja a -1			1	,2

aproxima a la normalidad en el caso de la distribución asimétrica. No obstante, hemos de indicar que los valores más pequeños son los obtenidos con el índice Rwg*.

Finalmente, los valores de la kurtosis para el R_{wg} estuvieron por encima del valor recomendado por DeCarlo (1997) inferior a 3 (14,579 para la uniforme y 13,03 para la asimétrica). Por el contrario, los valores de la kurtosis fueron mínimos para el índice Rwg* (0,49 para las distribuciones uniforme y asimétrica).

En suma, tomados en conjunto estos resultados podemos apoyar la sugerencia realizada por Lindell et al. (1999) y Eby et al (1999) de que el índice Rwg* es más adecuado que el índice Rwg de James et al (1984) para reflejar la dispersión intragrupal en escalas con múltiples ítems.

Segundo estudio

Método

Muestra

Un total de 258 sujetos pertenecientes a 51 equipos de trabajo configuraron la muestra del segundo estudio. El promedio de edad de los participantes fue de 35,69 años (sd= 9,04), de los cuales 51,20% fueron hombres y el resto mujeres. El porcentaje de sujetos que desarrollan su actividad en la empresa pública fue del 51,4% y el 48,6% lo hacían en la empresa privada. La media de antigüedad en el equipo de trabajo se situó en 8,60 (sd= 8,3) y la media de antigüedad en la empresa fue de 11,48 (sd= 9,26). El 48,8 % de los sujetos tenían título universitario.

Procedimiento

Al igual que en el primer estudio los datos se recogieron utilizando el cuestionario de conducta organizacional (C.N.O.-2). A continuación examinamos en qué medida las percepciones de las conductas culturales de los sujetos podrían ser extrapolables a nivel grupal, por lo que nos servimos de los índices de acuerdo de James et al. y de Lindell et al. Para poder comparar estos dos índices se siguió el procedimiento utilizado en el primer estudio.

Resultados

En la tabla 3 se presentan los valores obtenidos utilizando los dos índices de acuerdo, el de James et al. (1984) y el de Lindell et al. (1999). Como podemos observar, casi todos los valores Rwg están en los rangos más altos, mientras que los valores Rwg* muestran considerablemente una mayor variabilidad.

Tabla 2
Distribución de los índices de acuerdo

	Rwg* Uniforme	Rwg* Asimétrica	Rwg Asimétrica	Rwg Uniforme
equipos	39	39	39	39
Parámetros normales	Media	,4805	,2240	,8764
	Desviación típica	,1772	,2643	,6145
Z		1,077	1,092	1,168
Sig.		,196	,184	,131

La correlación entre los índices de acuerdo y su correspondiente varianza promedio también puso de manifiesto la superioridad del índice de acuerdo Rwg* (tabla 4). La correlación entre los valores obtenidos utilizando el índice de James et al. (Rwg) y su correspondiente varianza promedio sólo resultó significativa en el caso de la distribución asimétrica (p<,027), sin embargo cuando se utilizaron los valores obtenidos por medio del índice de Lindell et al. (Rwg*) resultó una correlación perfecta con la varianza promedio (p<,000). Una vez más se pone de manifiesto que el índice Rwg* se corresponde más fielmente a las diferencias en la varianza media de las respuestas dentro del grupo.

Respecto a la prueba de la normalidad, el estadístico de Kolmogorov-Smirnov sólo se aproxima a la normalidad utilizando el índice Rwg*, tanto en el supuesto de la distribución uniforme (z= ,645; p<,800) como asimétrica (z=,643; p<,800), pero no así cuando se utiliza el índice Rwg (z= 2,016, p <,001; z= 3,828, p <,000). Utilizando la prueba de la kurtosis encontramos que los valores encontrados utilizando el índice de acuerdo de Lindell et al. son inferiores a tres (valor recomendado por DeCarlo, 1997), en concreto 1,033 tanto en el supuesto de una distribución asimétrica como uniforme. Por el contrario, cuando utilizamos el índice de James et al. la kurtosis alcanzó valores de 6,881 y 51, respectivamente.

En suma, tomados en conjunto estos resultados podemos afirmar que en este segundo estudio se demuestra con mayor fuerza la sugerencia realizada por Lindell et al (1999) y Eby et al. (1999) de que el índice Rwg* es más válido y adecuado para evaluar el

Tabla 3
Variabilidad de los índices de James et al (1984) y Lindell et al (1999)

Valores	Rwg* Uniforme	%	Rwg* Asimétrica	%	Rwg Uniforme	%	Rwg Asimétrica	%
De 1,01 al más alto					21	2,1	93	15,4
De ,9 a 1	8	1,3	5	,9	296	29,8	160	26,4
de ,8 a ,89	73	12,2	20	3,4	129	13,0	122	20,2
de ,7 a ,79	84	14,0	56	9,6	72	7,3	55	9,1
de ,6 a ,69	100	16,7	65	11,2	32	3,2	32	5,3
de ,5 a ,59	74	12,3	55	9,5	19	1,9	27	4,5
de ,4 a ,49	68	11,3	64	11,0	7	,7	18	3,0
de ,3 a ,39	69	11,5	51	8,8	8	,8	11	1,8
de ,2 a ,29	47	7,8	45	7,7	4	,4	13	2,1
de ,1 a ,19	31	5,2	47	8,1	1	,1	8	1,3
de ,0 a ,09	19	3,2	52	8,9	5	,5	9	1,5
de -,1 a -,19	5	,8	32	5,5	4	,4	6	1,0
de -,2 a -,29	5	,8	25	4,3	1	,1	1	,2
de -,3 a -,39	3	,5	14	2,4	1	,1	6	1,0
de -,4 a -,49	3	,5	12	2,1			1	,2
de -,5 a -,59	2	,3	7	1,2			2	,3
de -,6 a -,69	1	,2	7	1,2			1	,2
de -,7 a -,79	4	,7	4	,7			3	,5
de -,8 a -,89	0	0	4	,7			1	,2
de -,9 a -,99	1	,1	2	,3			1	,2
de la más baja a -1	3	,3	1,5	2,6	10	1,0	35	5,8

Tabla 4
Correlaciones entre la varianza y los índices de acuerdo

	Varianza	Rwg Asimétrica	Rwg* Asimétrica	Rwg Normal	Rwg* Normal
VARIANZA	1,000	-,309	-1,000	,173	-1,000
Sig. (bilateral)		,027	,000	,224	,000
Nº de grupos	51	51	51	51	51

acuerdo intragrupal en escalas con múltiples ítems que el propuesto por James et al. (1984).

Discusión

El objetivo de esta investigación ha sido desarrollar un sistema de análisis de los datos procedentes de los sujetos para poder ser extrapolables a nivel de análisis grupal; en nuestro caso concreto hemos evaluado las percepciones que tienen los sujetos de las normas culturales para poder determinar en que grado existe una cultura de equipo de trabajo. Como el nivel correspondiente de análisis es el grupal se han comparado dos índices, el de James et al. (1984) y el de Lindell et al. (1999), que son los que se han venido utilizando en la literatura para evaluar el acuerdo intragrupal cuando se trata de medidas de agregación. Los resultados de este estudio ponen de relieve que el índice Rwg* para analizar el acuerdo intragrupal es una técnica viable para evaluar las normas compartidas a nivel de grupo. Normas grupales que son referidas a los patrones de conducta regulares que son relativamente estables dentro de un grupo (Jackson, 1965).

Tanto el estudio 1, como el estudio 2, sugieren que el estadístico Rwg* es sensible a las diferencias intra-grupales, es decir, respecto al grado en que se comparten estas normas grupales. En ambos estudios se ha examinado la relación entre índices de acuerdo intragrupal (Rwg*, Rwg) y la varianza media del grupo, proporcionando evidencia de la superioridad del estadístico Rwg* como un índice de convergencia perceptiva entre los miembros del equipo. Por otro lado, el Rwg* (pero no el Rwg) presenta una distribución normal. Estos hallazgos suponen una contribución importante a los métodos de investigación organizacional y grupal ya que los investigadores se han esforzado en averiguar cómo medir el contenido de modelos mentales a diferentes niveles de análisis, señalando la particular dificultad asociada con evaluar el acuerdo dentro de los grupos. Nuestro estudio presenta un valor añadido más, si cabe, y es la confirmación de los resultados del estudio de Eby et al. (1999), pero en nuestro caso concreto con grupos o equipos de trabajo reales y no pseudo-equipos (por ejemplo estudiantes universitarios que forman grupos para realizar unas determinadas tareas).

Nuestra confirmación de la superioridad del índice Rwg* sobre el índice Rwg también tiene implicaciones metodológicas para la investigación en equipos. En concreto, si los investigadores están interesados en utilizar el acuerdo intragrupal como un indicador de fenómenos compartidos a nivel de grupo (léase p.ej., expectativas compartidas, significados compartidos...), será más adecuado que utilicen el Rwg* más que el Rwg.

Nuestro trabajo también presenta algunas implicaciones para futuras investigaciones y desarrollos teóricos. Por ejemplo, una ampliación de este estudio podría ser comparar diferentes métodos de evaluar el consenso o el acuerdo donde se varíen las características de las puntuaciones de las escalas y la naturaleza de las respuestas de los miembros. Esto parece particularmente importante dado que –como indicaron Bliese y Halverson (1998)– la relación entre consenso y la puntuación promedio del grupo a un constructo aumenta cuando los miembros dirigen sus respuestas a los extremos de una escala. En esta línea, también sería de interés que trabajos posteriores comparasen diferentes métodos entre sí, por ejemplo el método del consenso y el método de agregación. En la literatura ya existe algún precedente. Por ejemplo, el trabajo de Kirkman et al., (2001). No obstante, estos autores se limitaron a comparar el método de consenso con el índice de James et al. (1984), llegando a la conclusión de que el método del consenso es mejor que el método de agregación. Los resultados de nuestro estudio, no obstante, ponen en cuestión este resultado, dado que no utilizaron el índice de Lindell et al. (1999) que, según nuestro estudio tiene una mayor validez predictiva que el índice de James et al.

Por otro lado, es importante señalar que los métodos basados en la agregación y en el consenso probablemente puedan presentar una sensibilidad diferencial a las fuentes de error en la medición. Por ello, investigaciones posteriores habrán de tener presente estas fuentes así como comparar diferentes fuentes de error y métodos de estimación de la fiabilidad cuando se utiliza la agregación, el consenso u otros métodos de evaluación de constructos al nivel de equipo. La utilización de diferentes métodos y la consideración de diferentes fuentes de error puede acarrear una mejor comprensión de los constructos al nivel de equipo y de su medición.

Referencias

- Alonso, E. (1997). *El método del calibrado en la evaluación de la cultura organizacional*. Trabajado presentado al VI Congreso Nacional de Psicología Social, San Sebastián
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Nueva York: Freeman
- Bliese, P.D. y Halverson, R.R. (1998). Group consensus and psychological well-being. A large field study. *Journal of Applied Social Psychology*, 28, 563-580.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83, 234-246.
- DeCarlo, L.T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2, 292-307.
- Earley, P.C. (1999). Playing follow the leader: Status-determining traits in relation to collective efficacy across cultures. *Organizational Behavior and Human Decision Processes*, 80, 192-212.
- Eby, L.T., Meade, A.W., Parisi, A.G. y Douthitt, S.S. (1999). The development of an individual-level teamwork expectations measure and the application of a within-group agreement statistic to assess shared expectations for teamwork. *Organizational research methods*, 2(4), 366-394.
- Feltz, D.L. y Ligg, C.D. (1998). Perceived team and player efficacy in hockey. *Journal of Applied Psychology*, 83, 557-564.
- Fleener, J.W., Fleener, J.B. y Grossnickle, W.F. (1996). Interrater reliability and agreement of performance ratings: A methodological comparison. *Journal of Business and Psychology*, 10, 367-380.
- Gibson, C.B., Randel, A.E. y Earley, A.E. (2000). Understanding group efficacy: An empirical test of multiple assessment methods. *Group and Organization Management*, 25, 67-97.
- Gist, M.E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review*, 12, 472-485.
- Jackson, J. (1965). Structural characteristics of norms. En I.D. Steiner y M. Fishbein (Eds.), *Current studies in social psychology*. Nueva York: Holt, Rinehart and Winston.
- James, L.R., Demaree, R.J. y Wolf, G. (1984). Estimating within group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85-98.

- James, L.R., Demaree, R.J. y Wolf, G. (1993). r'_{wg} . An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 28, 306-309.
- Kirkman, B.L., Tesluk, P.E. y Rosen, B. (2001). Assessing the incremental validity of team consensus ratings over aggregation of individual level data in predicting team effectiveness. *Personnel Psychology*, 54, 645-667.
- Klein, K.J., Conn, A.B. Smith, D.B. y Sorra, J.S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology*, 86, 1, 3-16.
- Lindell, M.K. (2001). Assessing and testing interrater agreement on a single target using multi-item rating scales. *Applied Psychological Measurement*, 25(1), 89-99.
- Lindell, M.K. y Brandt, C.J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the *CVI*, *T*, *Rwg(j)*, and *R*wg* indexes. *Journal of Applied Psychology*, 84, 640-647.
- Lindell, M.K., Brandt, C.J. y Whitney, D.J. (1999). A revised index of inter-rater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, 23, 127-135.
- Lindell, M.K. y Brandt, C.J. (1997). Measuring inter-rater agreement for ratings of a single target. *Applied Psychological Measurement*, 21, 271-278.
- Lindsley, D.H., Brass, D.J. y Thomas, J.B. (1995). Efficacy-performance spirals: A multilevel perspective. *Academy of Management Review*, 20, 645-678.
- Nunnally, J.C. (1978). *Psychometric Theory*. Nueva York: McGraw-Hill.
- Sánchez, J.C. y Alonso, E. (1997). *La evaluación de la cultura: El cuestionario de conductas normativas organizacionales*. Trabajo presentado al VI Congreso Nacional de Psicología Social, San Sebastián.
- Sánchez, J.C. y Alonso, E. (1998). *Análisis de las escalas del cuestionario de conductas normativas (CNO-2)*. Trabajo presentado al IV Congreso Nacional de Psicología del Trabajo y de las Organizaciones. Valladolid.
- Strelau, J., Angleitner, A., Bantelmann, J. y Ruch, W. (1990). The Strelau temperament inventory-revised (STI-R): Theoretical considerations and scale development. *European Journal of Personality*, 4, 209-235.