ARTICLE IN PRESS

Journal of Experimental Social Psychology xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

FlashReport

A nonparametric method to analyze interactions: The adjusted rank transform test

Christophe Leys *, Sandy Schumann

Université Libre de Bruxelles, Social Psychology Unit CP 122, Avenue F. Roosevelt, 50, 1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 18 December 2009
Revised 21 February 2010
Available online xxxx

Keywords:
Nonparametric test
Factorial design
Interaction
Rank test
Adjusted rank transformation

ABSTRACT

Experimental social psychologists routinely rely on ANOVA to study interactions between factors even when the assumptions underlying the use of parametric tests are not met. Alternative nonparametric methods are often relatively difficult to conduct, have seldom been presented into detail in regular curriculum and have the reputation – sometimes incorrectly – of being less powerful than parametric tests. This article presents the adjusted rank transform test (ART); a nonparametric test, easy to conduct, having the advantage of being much more powerful than parametric tests when certain assumptions underlying the use of these tests are violated. To specify the conditions under which the adjusted rank transform test is superior to the usual parametric tests, results of a Monte Carlo simulation are presented.

Introduction

Psychological researchers often aim at investigating the impact of one or more factors on an outcome variable. Rather than conducting several studies, each studying the impact of a single factor, researchers often want to evaluate the joint impact of these factors and study their relevant interactions. As an example, social psychologists could suspect an impact of a defendant's emotional behavior on the outcome of the sentence carried out by a judge in a trial court (Leys, Licata, & Klein, 2010). Anger and Guilt could be two pertinent emotions to set as independent variables while the trial's sentence would be the dependent variable. Incorporating two IVs is especially preferred when their interaction is relevant for the research question. This would be the case if the presence (or absence) of guilt impacts differently on the sentence whether the defendant displays anger or does not.

The statistical approach for analyzing such a design is the parametric factorial ANOVA (Keppel, 1991; Sheskin, 2004). But, the validity of this test can be jeopardized when its underlying assumptions are not met (Sawilowsky, 1990). In that case, conducting nonparametric tests becomes an interesting option.

In social psychology, nonparametric tests receive little attention in the regular curriculum (Buday & Kerr, 2000). The literature that is most frequently referred to (Howell, 1999; Kenny, Kashy, & Bolger, 1998; Keppel, 1991) deals mostly with common nonparametric tests but not with complex issues, like interactions, leaving researchers confronted with a conundrum. Two reviews support this assumption: Sherman, Buddie, Dragan, End, and Finney (1999) analyzed all publications within 20 years in the *Personality and Social Psychology Bulletin* (*PSPB*) whereas Stone-Romero, Weaver, and Glenar (1995) analyzed the *Journal of Applied Psychology (JAP).* Fifteen percents of all studies in PSPB and less than height in JAP used nonparametric tests. Furthermore, we did a quick review of those articles and found no studies at all using nonparametric test to study interaction.

Nonetheless, we argue that in some violations of ANOVA's assumptions, that we will detail further, nonparametric tests yield more robust results. In view of the paucity of alternatives to ANOVAs, this article presents a nonparametric method to analyze interactions – the adjusted rank transform test (ART) – which has the advantage of being easy to conduct, offers an interesting robustness, does not depend on the distribution of the variables and is based on regular *F* distribution tables. In addition, it can be applied to within-subjects, between-subjects or mixed experimental designs.

Requirements for conducting a factorial ANOVA

To conduct a factorial ANOVA it is necessary that the distribution of the dependant variables in the population is normal which is inferred from the sample's distribution that, therefore, has to be close to a normal distribution as well. The variance must be comparable between the different experimental conditions (homoscedastic). The sampling must be simple random. Lastly, the dependant variable is required to be at least measured on an interval scale.

Consequences of violations of the assumptions

Whenever one or more of these conditions are violated an increase in type I or type II error may occur. Type I error is the prob-

* Corresponding author.
E-mail address: cleys@ulb.ac.be (C. Leys).

0022-1031/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2010.02.007

ability of falsely rejecting the null hypothesis. The type II error is the probability of falsely accepting the null hypothesis. The power of a test is the probability to avoid the type II error.

Arguments supporting the use of parametric tests

Violation of factorial ANOVA does not exclude its use: it is, for example, possible to apply Likert scales and still use parametric tests without major consequences (Nanna & Sawilowsky, 1998; Zumbo & Zimmerman, 1993); A deviation from a normal distribution when the sample size is above 30 participants per condition yield robust results (Glass, Peckham, & Sanders, 1972; Lumley, Diehr, Emerson, & Chen, 2002); When homoscedasticity is violated, a correction implying a loss in degrees of freedom or a transformation of the raw data like using the logarithm or inverting the scores is acceptable (Howell, 1999; Keppel, 1991; Lix, Keselman, & Keselman, 1996). Besides, most nonparametric tests are not immune to violations of homosedasticity (Tomarken & Serlin, 1986) and thus can not be used as an alternative.

Arguments supporting a shift towards nonparametric tests

When several violations of assumptions occur, the power, and the type I error, of a parametric test is often reduced, and therefore requires a shift towards nonparametric options (Sawilowsky, 1990). For example, if the samples in different experimental conditions are not of the same size AND the variances are heterogeneous, the power of the test is reduced drastically. This is especially true if the experimental condition with the biggest sample size is the one with the smallest variance (Box, 1953; Snedecor & Cochran, 1980). As we will see, using the ART is advisable if the sample size is under 30 per experimental condition and the requirement of a normal distribution is not fulfilled, or when heteroscedacity occurs along with a non-normal distribution.

Nonparametric alternatives to factorial ANOVA

Many options are available to study interactions but either the procedure is very complex or is lacking in power. Among the complex solutions are log-linear analysis (Christensen, 1997; Klein & Azzi, 1999); the Bray-Curtis ordination (Anderson, 2001); the Welsh-James test (Keselman, Algina, Wilcox, & Kowa, 2000); the adapted L statistic of rank test (Puri & Sen, 1971, 1985).

Easier to conduct, the ART is based on the test of rank transformation (RT) introduced by Conover and Iman (1981). The guiding principle of RT is to assign a rank to each given raw data according to a classical method – that is described later – and to conduct a regular parametric test on those ranks. Ranking the data allows them to be distribution free. Sawilowsky (1990) reports a RT's statistical power three times higher than a factorial ANOVA (.91 instead of .32) under non-normal distribution.

However, the presence of both the main effects and significant interactions alter the robustness of the test. It is therefore necessary to incorporate an adjustment that allows us to avoid this inconvenience.

How to adjust the RT into ART to study interaction?

Sawilowsky (1990) describes an adjustment of the RT by removing the main effects according to the principles that were proposed by Hettmansperger and McKean (1978): an interaction is defined mathematically as the remaining effect after having deducted all other effects that might contribute to the overall mean of a group or experimental condition. The "other effects" are the main effects of the factors, the general mean and the subjects' residual error. The main effect of a factor is the effect of one factor without taking into account the effect of the other factor.

Since the grand mean is supposed to be the same for all samples and the individual residuals are assumed to be distributed in a uniform way within each condition, only the main effect has to be removed. This can be done by subtracting the respective marginal means from each observation(Rosnow & Rosenthal, 1995). Afterwards, a rank is assigned to each observation pooled together and a factorial ANOVA is conducted. The significance of the interaction term is interpreted just like in the classical ANOVA. Using this procedure, information concerning the main effects is necessarily lost. However, they can be computed in a second step by deducting the interaction of the raw data keeping only the main effects. An example for this calculation is provided below.

Computer simulations

To define the conditions under which ART is a preferable alternative to parametric tests, we performed simulations using the Monte Carlo method for different hypothetical distributions in which the assumptions underlying ANOVA are not fulfilled (Leys, 2009).

Distribution other than normal

Considering a random sample of 20 subjects per conditions, in a 2×2 design, the ART is much more powerful than the parametric test when the condition of normal distribution is not fulfilled. This was tested with alternative non-normal distributions likely to be observed in psychological research such as truncated normal and student distributions and also for asymmetric distributions such as Gamma.

Heterosedasticity

Confirming previous studies (Tomarken & Serlin, 1986), both the parametric and the nonparametric tests are sensitive to heterosedasticity. Yet, the implications for the analysis of an interaction and main effects are divergent. For both, the ART presents the risk of a slightly inflated alpha; i.e., when there is no effect, the test detects one in approximately 6% of the cases if the alpha is set at 5%. In a few cases the likelihood of type I error increases up to 9% for the analysis of the main effects when using the adjusted rank transform test whereas it does not get over 7% for the analysis of interaction. The factorial ANOVA never reaches 7% of type I error inflation.

Analyzing interaction with the ART in heterosedastic conditions yields less powerful results than the classic factorial ANOVA whereas analyzing main effects with ART is more powerful. Nevertheless, nonparametric tests present the additional disadvantage of varying greatly in power: some combinations of the main effects and interactions can lead to a drastic loss of power whereas other combinations result in a very satisfying power. Hence, the wisest choice in case of heterosedasticity, as the sole violated condition, is to opt for a more complex method like the Welsh-James test, which is immune to heterosedasticity, or to adapt the data, as proposed before, in order to perform a factorial ANOVA.

Heterosedasticity and non-normal distributions

Finally, if both normality and homosedasticity are violated, the ART is much more powerful then the factorial ANOVA even for slight deviations. This is true for the analysis of interactions or main effects. The difference in power between factorial ANOVA and ART increases as the deviations from normality and the heterosedasticity get larger.

Please cite this article in press as: Leys, C., & Schumann, S. A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology* (2010), doi:10.1016/j.jesp.2010.02.007

Table 1
Raw data of the fictive case.

	A_1	A_0	Marginal mean
B_1	6	2	
	1	3	
	7	2	$\overline{B_1} = 4.00$
	8	1	
	9	1	
Group mean	$\overline{AB_{11}} = 6.20$	$\overline{AB_{01}} = 1.80$	
B_0	1	1	
	2	3	
	3	2	$\overline{B_0} = 3.40$
	3	7	
	4	8	
Group mean	$\overline{AB_{10}} = 2.60$	$\overline{AB_{00}} = 4.20$	
Marginal mean	$\overline{A_1} = 4.40$	$\overline{A_0} = 3.00$	$\overline{X} = 3.70$

Table 2
Deduction of the marginal mean of the line and column from the observed data.

	A_1	A_0	Marginal mean
B_1	6–4–4.4 = **–2.4**	2–4–3 = **–5**	
	1–4–4.4 = **–7.4**	3–4 – 3 = **–4**	
	7–4–4.4 = **–1.4**	2–4–3 = **–5**	$\overline{B_1} = 4.00$
	8–4–4.4 = **–0.4**	1–4–3 = **–6**	
	9–4–4.4 = **0.6**	1–4–3 = **–6**	
Group mean	$\overline{AB_{11}} = 6.20$	$\overline{AB_{01}} = 1.80$	
B_0	1–3.4–4.4 = **–6.8**	1–3.4–3 = **–5.4**	
	2–3.4–4.4 = **–5.8**	3–3.4–3 = **–3.4**	
	3–3.4–4.4 = **–4.8**	2–3.4–3 = **–4.4**	$\overline{B_0} = 3.40$
	3–3.4–4.4 = **–4.8**	7–3.4–3 = **0.6**	
	4–3.4–4.4 = **–3.8**	8–3.4–3 = **1.6**	
Group mean	$\overline{AB_{10}} = 2.60$	$\overline{AB_{00}} = 4.20$	
Marginal mean	$\overline{A_1} = 4.40$	$\overline{A_0} = 3.00$	$\overline{X} = 3.70$

A numerical example to conduct the ART

To understand the ART better, a set of fictional raw data simulate the example of a 2 × 2 design presented in introduction (see Table 1): The severity of the sentence that should be inflicted on an offender (1 = very harsh, 9 = very mild) is the DV. The two IVs are guilt as a factor *A* (i.e., whether this offender expressed a feeling of guilt, A_1, or not, A_0, during the trial) and anger as a factor *B* (i.e., whether the offender expresses anger, B_1, or not, B_0). The marginal means suggest that expressing guilt or expressing anger could yield milder sentences. Yet, looking at the groups means suggest an interaction: the effect of the expression of guilt might be moderated by the expression of anger.

Studying interaction

The data are adjusted by subtracting the sum of the marginal mean of the line and the column from each relevant observation. This isolate the interaction by removing the main effects (see Table 2).

Ranks are then assigned to the pooled adjusted observations (see Table 3) by: (a) aligning all the observations in an increasing order; (b) assigning a rank ranging from one to *N* in an increasing order and (c) adapting the rank for tie values by assigning the same averaged rank to each one.[1] For example, the score "−6" appears twice and the scores are labeled with ranks three and four. In this case, the rank for both of the original scores of "−6" is the mean of three and four – namely 3.5.

Lastly the factorial ANOVA on the adjusted ranked data is conducted (see Tables 4 and 5). As expected, we obtain a significant interaction ($p = .02$) meaning that the impact of guilt on the sentence is significantly different when the defender feels angry than when he does not. The effect or the partial effect sizes (η^2) are easy to compute just like in a classical factorial ANOVA (see Table 5).

Studying main effects

Main effects can be isolated by subtracting the interaction from the raw data (see Table 6). The interaction is deducted by subtracting the mean of the two diagonal group means from each observation. For example, the score 5.20 ($=\frac{6.20+4.20}{2}$) is deducted from the observations in condition A_1B_1 and B_0A_0. In the same way, the score 2.20 is deducted from the observations in the conditions A_0B_1 and A_1B_0. As a next step, a parametric test follows where the main effects are calculated without being affected by the interaction (Table 7). Results show a marginally significant main effect of the factor *A*, guilt which, as in regular ANOVA, should not be straightforwardly interpreted since the significant interaction shows a moderation between anger and guilt.

Studying simple effects

A simple effect encompasses both the main effect and the interaction of one factor on a dependant variable in one modality of an-

[1] To rank the variables, use the syntax command RANK VARIABLES with SPSS and PROC RANK with SAS.

Table 3
Transforming the adjusted raw data to ranks.

Observations	−7.4	−6.8	−6	−6	−5.8	−5.4	−5	−5	−4.8	−4.8
Number	1	2	3	4	5	6	7	8	9	10
Rank	**1**	**2**	**3.5**	**3.5**	**5**	**6**	**7.5**	**7.5**	**9.5**	**9.5**
Observations	−4.4	−4	−3.8	−3.4	−2.4	−1.4	−0.4	0.6	0.6	1.6
Number	11	12	13	14	15	16	17	18	19	20
Rank	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18.5**	**18.5**	**20**

Table 4
Replacing the adjusted scores with ranks.

	A_1	A_0	Marginal mean
	15	7.5	
	1	12	
B_1	16	7.5	$\overline{B_1} = 10.15$
	17	3.5	
	18.5	3.5	
Group mean	$\overline{AB_{11}} = 13.5$	$\overline{AB_{01}} = 6.80$	
	2	6	
	5	14	
B_0	9.5	11	$\overline{B_0} = 10.85$
	9.5	18.5	
	13	20	
Group mean	$\overline{AB_{10}} = 7.80$	$\overline{AB_{00}} = 13.9$	
Marginal mean	$\overline{A_1} = 10.65$	$\overline{A_0} = 10.35$	$\overline{X} = 10.5$

Table 5
Factorial ANOVA to test the interaction.

Source	*SS*	*df*	*F*	η^2	*p*
Interaction $A \times B$	204.80	1	7.20	.31	.02
Error	455.30	16			
Total	2868.00	20			

Note: Being irrelevant, main effects are not presented.

Please cite this article in press as: Leys, C., & Schumann, S. A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology* (2010), doi:10.1016/j.jesp.2010.02.007

Table 6
Scores of the fictive example from which the interaction is deducted.

	A_1		A_0		Marginal mean
	Observations	Rank	Observations	Rank	
	6–5.20 = .80	12	2–2.20 = −.20	9	
	1–5.20 = −1.20	5.5	3–2.20=.80	12	
B_1	7–5.20 = 1.80	16	2–2.20 = −.20	9	$\overline{B_1} = 4.00$
	8–5.20 = 2.80	18.5	1–2.20 = −1.20	5.5	
	9–5.20 = 3.80	20	1–2.20 = −1.20	5.5	
Group mean	$\overline{AB_{11}} = 6.20$		$\overline{AB_{01}} = 1.80$		
	1–2.20 = −1.20	5.5	1–5.20 = −4.20	1	
	2–2.20 = −.20	9	3–5.20 = −2.20	3	
B_0	3–2.20=.80	12	2–5.20 = −3.20	2	$\overline{B_0} = 3.40$
	3–2.20=.80	12	7–5.20 = 1.80	16	
	4–2.20 = 1.80	16	8–5.20 = 2.80	18.5	
Group mean	$\overline{AB_{10}} = 2.60$		$\overline{AB_{00}} = 4.20$		
Marginal mean	$\overline{A_1} = 4.40$		$\overline{A_0} = 3.00$		$\overline{X} = 3.70$

Table 7
Overall table of the factorial ANOVA using ART.

Sums of squares of ART with rank adjusted to remove main effects			
SS interaction $A \times B$	204.80		
SS error	455.30		
Sums of squares of ART with rank adjusted to remove interaction			
SS main effect *A*	101.25		
SS main effect *B*	16.20		
Error	511.40		
Sums of squares for simple effects with both adjustment, and SS reconstructed			
SS_iA_1/B	81.23	$SSA_1/B = \left(\sqrt{81.23} + \sqrt{30.63}\right)^2 = 211.62$	
$SS_{me}A_1/B$	30.63		
SS_iA_0/B	126.03	$SSA_0/B = \left(\sqrt{126.03} + \sqrt{.03}\right)^2 = 129.95$	
$SS_{me}A_0/B$	.03		
Best estimation of error sum of square by averaging both error sums of squares			
Average SS error	(455.30 + 511.40)/2 = 483.35		
Overall F (using average SS error)			
	F (*1, 16*)	*p*	
Interaction	6.78	.02	
Main effect *A*	3.35	.09	
Main effect *B*	.54	.47	
Simple effect A_1/B	7.00	.02	
Simple effect A_0/B	4.30	.05	

other factor. In our example, the two modalities of factor *B* will be compared in each of the modalities of factor *A* (Table 7). Here, $\overline{AB_{11}}$ will be compared with $\overline{AB_{10}}$ and $\overline{AB_{01}}$ with $\overline{AB_{00}}$.

Yet, since the main effect and the interaction have been computed separately, the sum of squares of the simple effects has to be reconstructed[2] using the following three steps:

(1) Compute the two simple effects of the interaction, as it would be done in a regular factorial ANOVA and retain the sum of squares of the simple effects. In our example we will note them as SS_iA_1/B and SS_iA_0/B.
(2) Compute the two simple effects of the main effects and retain the sum of squares of those simple effects noted as $SS_{me}A_1/B$ and $SS_{me}A_0/B$.
(3) Recreate the global simple effects sum of squares by squaring the sum of the square roots of both the main effect and the interaction simple effects: $SSA_1/B = \left(\sqrt{SS_{me}A_1/B} + \sqrt{SS_iA_1/B}\right)^2$ and identically for SSA_0/B.

[2] To help computing the ART, we created an Excel spreadsheet to rebuild the simple effects, compute the error term and display the overall *F*-ratio for the interaction, main effects and simple effects in a 2×2 factorial design: www.psycho-psysoc.site.ulb.ac.be/images/stories/file/ART%202×2%20factorial%20design.xls.

The last issue is to define the sum of squares of the error. It is, by definition, independent of any experimental treatment. Therefore, both error terms of the interaction and of the main effects ANOVA are relevant estimations of the same error. We recommend using the average of those two estimations (in fact we would even recommend using this computed error term in each ANOVA table we computed until now, as it is done in Table 7).

Thus, the results can be interpreted as follow: If the defendant feels guilty, then feeling angry will lead to a significantly milder sentence than feeling no anger. Conversely, if the defendant does not feel guilty, then feeling angry will lead to a significantly harsher sentence than feeling no anger.

Conclusion

When studying interactions, parametric analyses are not always advisable if several assumptions are violated. We have presented an alternative: the adjusted rank transform test which is a middle ground between parametric and nonparametric methods. It follows the nonparametric approach because it is based on a test of ranks but it is close to parametric tests since the *F* distribution tables are used. This method looses much of its robustness as soon as the main effects occur together with one or several interactions. To avoid this problem, the scores are adjusted by deducting the main effects or, the interaction, and then analyzing separately the interactions, or the main effects.

Acknowledgments

We would like to express our deepest gratitude to Dominique Muller and Daniel Holender for their precious counsels and expertise. We also thank Olivier Klein for his indefectible support and his genuine concern during the whole process of this publication.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jesp.2010.02.007.

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26*, 32–46.
Buday, J. F. E., & Kerr, D. (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology, 27*, 248–257.
Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika, 40*, 318–335.

Please cite this article in press as: Leys, C., & Schumann, S. A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology* (2010), doi:10.1016/j.jesp.2010.02.007

Christensen, R. (1997). *Log-linear models and logistic regression*. New-York: Springer.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician, 35*, 124–129.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*, 237–288.

Hettmansperger, T., & McKean, J. (1978). Statistical inference based on ranks. *Psychometrika, 43*, 69–79.

Howell, D. C. (1999). *Fundamental statistics for the behavioral sciences*. Belmont, CA: Thomson Wadsworth.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.). *Handbook of social psychology* (Vol. 233–265). Boston, MA: McGraw-Hill.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Engelwood Cliffs: J: Prentice-Hall.

Keselman, H. J., Algina, J., Wilcox, R. R., & Kowa, R. K. (2000). Testing repeated measures hypotheses when covariance matrices are heterogeneous: Revisiting the robustness of the Welch-James test again. *Educational and Psychological Measurement, 60*, 925–938.

Klein, O., & Azzi, A. (1999). L'analyse log-linéaire en psychologie sociale: Une introduction [Log-linear analysis in social psychology: An introduction]. *Cahiers Internationaux de Psychologie Sociale, 42*, 102–124.

Leys, C. J. (2009). A Monte Carlo analysis of the adjusted rank transform test applied in a two-by-two factorial design. Université Libre de Bruxelles. Unpublished manuscript.

Leys, C. J., Licata, L., & Klein, O., (2010). The influence of a defendant's emotional behaviour on the severity of the trial's sentence. Université Libre de Bruxelles. Unpublished manuscript.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the factorial analysis of variance "*F*" test. *Review of Educational Research, 66*, 579–619.

Lumley, T., Diehr, P., Emerson, S., & Chen, L., (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23, 151–169.

Nanna, M. J., & Sawilowsky, S. S. (1998). Analysis of Likert scale data in disability and medical rehabilitation research. *Psychological Methods, 3*, 55–67.

Puri, M. L., & Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. New-York: Wiley.

Puri, M. L., & Sen, P. K. (1985). *Nonparametric methods in general linear models*. New-York: Wiley.

Rosnow, R. L., & Rosenthal, R. (1995). Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science, 6*, 3–9.

Sawilowsky, S. S. (1990). Nonparametric tests of interaction in experimental design. *Review of Educational Research, 60*, 91–126.

Sherman, R. C., Buddie, A. M., Dragan, K. L., End, C. M., & Finney, L. J. (1999). Twenty years of PSPB: Trends in content, design, and analysis. *Personality and Social Psychology Bulletin, 25*, 177–187.

Sheskin, D. (2004). *Handbook of parametric and nonparametric statistical procedures*. Florida: CRC Press.

Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods*. Iowa: Iowa State University Press.

Stone-Romero, E. F., Weaver, A. E., & Glenar, J. L. (1995). Trends in research design and data analytic strategies in organizational research. *Journal of Management, 21*, 141.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin, 99*, 90–99.

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology, 34*, 390.

Please cite this article in press as: Leys, C., & Schumann, S. A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology* (2010), doi:10.1016/j.jesp.2010.02.007