Psicológica (2014), 35, 361-383.

Investigating the robustness of the nonparametric Levene test with more than two groups

David W. Nordstokke* and S. Mitchell Colp

University of Calgary, Canada

Testing the equality of variances during hypothesis testing is an important preliminary step before using statistical tests such as the t-test or ANOVA. It has been demonstrated that many tests for equality of variances are sensitive to non-normal distributions. Using computer simulation, the present simulation study investigates the Type I error rate and statistical power of the nonparametric and median versions of the Levene test for equality of variances when there are three, four or five groups used in the analysis. For each of the three, four and five group conditions there are several levels of sample size, variance ratio, group sample size imbalance, and degree of skew in the population distribution included in the simulation. Results show that the nonparametric Levene test shows good statistical properties when samples come from heavily skewed population distributions, when overall sample size was small, and when groups were unbalanced. The findings also allow for a relative comparison of the median-based Levene test of equality of variances under a variety of conditions. Practical implications for the testing for equality of variances are discussed.

A common practice in statistical data analysis in the psychological, behavioral and educational research is the comparison of means from two or more groups using an analysis of variance (ANOVA) type statistical test. A typical step in this accepted statistical practice is to conduct a test of equality of variances prior to the running of the ANOVA to determine whether or not the assumption of homogeneity of variances is tenable. Heterogeneity of variance occurs in when one or more groups of sample scores have a wider dispersion of scores than other groups to be used in a

* Address correspondence to: David W. Nordstokke, Ph.D. Educational Studies in School Psychology. Werklund School of Education. University of Calgary. 2500 University Drive NW. Education Tower 302. Calgary, Alberta, Canada T2N 1N4. Email: dnordsto@ucalgary.ca

between groups analysis. The consequence of heterogeneity of variances is that each group will contribute differentially to the estimation of the within groups variance parameter, and thus the sums of squares within groups will be a biased estimate of the population variance parameter leading to increases in the frequency of Type I and Type II errors and thus impacting the power of the test being used by reducing its capacity to correctly reject the null hypothesis.

Box (1953) noted that the F-test for equality of variances was overly sensitive in terms of inflated Type I error rates when the data distributions were sampled from non-normal (i.e., highly skewed or kurtotic distributions). Subsequent to Box's work, numerous tests of equality of variances have been developed (e.g., Levene, 1960; Brown & Forsythe, 1974). These tests were developed to be more robust to the violations of the assumptions of normality. Often these procedures involved transforming the raw score and carrying out an ANOVA on the transformed score. For example, the mean based Levene test transforms scores on the dependent variable by subtracting the mean from each score. Subsequent to this step, a one-way ANOVA is conducted using the transformed scores.

A nonparametric Levene (NPL) test was introduced by Nordstokke and Zumbo (2007) and has been shown to have good statistical properties in both simulated and real data settings (Nordstokke & Zumbo, 2010; Nordstokke, Zumbo, Cairns, & Saklofske, 2011). The NPL was developed as an extension of the mean based Levene where a rank transformation is applied to the data prior to conducting the ANOVA. This equates to using a parametric ANOVA on rank transformed data. The utilization of rank based transformations to avoid the assumption of normality was suggested by Friedman (1937) and more recently by Conover and Iman (1981) as a viable solution to nonnormal distributions. Statisticians and researchers generally agree that replacement of scores on the dependent variable by ranks before performing a parametric analysis of location yields the same decision as a nonparametric test (Zimmerman, 2012). The utilization of this approach is what gives the NPL its strengths for use in practical data analysis settings where data may come from nonnormal population distributions as the rank transformation reduces the impact of nonnormal data and outliers (Friedman, 1937).

As Nordstokke and Zumbo (2007; 2010) describe it, the steps of the NPL involves pooling the data from all groups, ranking the scores allowing, if necessary, for ties, placing the rank values back into their original groups, and running the Levene test on the ranks. The NPL test can be written as

$$\text{ANOVA}\ (\left|R_{ij} - \bar{X}_j^R\right|), \tag{1}$$

which is a one-way analysis of variance that is conducted on the absolute value of the mean of the ranks for each group, denoted $\bar{X}_j^R$, subtracted from each individual's rank R_{ij}, for individual i in group j. SPSS syntax used to compute the NPL for this study is listed in Appendix 1.

The purpose of the current study is to extend the simulation findings from Nordstokke and Zumbo (2010) to the three, four and five group ANOVA cases. To be consistent with Nordstokke and Zumbo (2010), the simulation includes results of the test that is often considered the "gold standard" of tests for equal variances, the Levene median (ML) test developed by Brown and Forsythe (1974) because, as Conover, Johnson and Johnson (1981) showed, one of the top performing tests for equality of variances in their simulation that compared 56 tests for equality of variances was the median based Levene test. The ML test for equality of variances can be expressed as,

$$\text{ANOVA}\ (\left|X_{ij} - Mdn_j\right|),$$

wherein, the analysis of variance is conducted on the absolute deviations of an individual's score, denoted X_{ij}, from their groups median value, denoted Mdn_j, for each individual i in group j.

This study will investigate the Type I error rates and statistical power of the NPL and ML in the three, four and five group ANOVA cases across several overall sample sizes with varying degrees of skew present in the population distribution, group imbalance and variance imbalance. The purpose of using a wide variety of conditions is to attempt to simulate a wide variety of conditions that might be found across a wide variety of research settings.

METHOD

Data Generation

Standard simulation methodology was employed to perform a computer simulation (e.g., Nordstokke & Zumbo, 2007; 2010; Zimmerman, 1987; 2004). Population distributions were generated and the statistical tests were performed using the statistical software package for the social sciences, SPSS 20. A pseudo random number sampling method with the initial seed selected randomly was used to produce χ^2 distributions. An example of the syntax used to create the population distribution of one group belonging to a normal distribution can be found in Appendix 1 of Nordstokke and Zumbo (2010). Building from Nordstokke and Zumbo (2007; 2010), the design of the three group simulation study was a 4 x 3 x 5 x 9 completely crossed design with: (a) four levels of skew of the population distribution, (b) three levels of sample size, (c) five levels of sample size ratio, $n1/n2/n3$, and (d) nine levels of ratios of variances. The dependent variables in this part of the simulation design are the proportion of rejections of the null hypothesis in each cell of the design and, more specifically, the Type I error rates (when the variances are equal), and power under the eight conditions of unequal variances. The design of the four group simulation study was a 4 x 3 x 7 x 7 completely crossed design with: (a) four levels of skew of the population distribution, (b) three levels of sample size, (c) seven levels of sample size ratio, $n1/n2/n3/n4$, and (d) seven levels of ratios of variances. Again, the dependent variables in this section of the simulation design are once again the proportion of rejections of the null hypothesis in each cell of the design and, more specifically, the Type I error rates (when the variances are equal), and power under the six conditions of unequal variances. The design for the five group simulation study was 4 x 3 x 3 x 5 completely crossed design with (a) four levels of skew, (b) three levels of sample size, (c) three levels of sample size ratio, $n1/n2/n3/n4/n5$, and (d) five levels of variance ratio.

Staying consistent with Nordstokke and Zumbo (2007; 2010), we only investigate and discuss statistical power in those conditions wherein the nominal Type I error rate, in our study .05(±.025), is maintained.

Shape of the population distributions[1]

Four levels of skew 0, 1, 2, and 3 were investigated. As is well known, as the degrees of freedom of a χ^2 distribution increase it more closely approximates a normal distribution. The skew of the distributions for both groups were always the same for every replication.

---

[1] It should be noted that the population skew was determined empirically for large sample sizes of 120,000 values with 1000, 7.4, 2.2, and .83 degrees of freedom resulting in skew values of 0.03, 1.03, 1.92, and 3.06, respectively; because the degrees of freedom are not whole numbers, the distributions are approximations. The mathematical relation is $\gamma_1 = \sqrt{8/df}$.

Sample Sizes

For the three group simulation, three different sample sizes, $N = n_1 + n_2 + n_3$, were investigated: 30, 60, and 90. Five levels of ratio of group sizes ($n_1 / n_2 / n_3$: 1/1/1, 1/1/4, 1/2/3, 3/2/1, and 4/1/1) were investigated. For the four group simulation, three different sample sizes, $N = n_1 + n_2 + n_3 + n_4$, were investigated: 40, 80, and 120. Seven levels of ratio of group sizes, ($n_1 / n_2 / n_3 / n_4$:1/1/1/1, 1/1/4/4, 1/1/2/4, 1/1/1/2, 2/1/1/1, 4/2/1/1, 4/4/1/1) were investigated. For the five group simulations, three different sample sizes, $N = n_1 + n_2 + n_3 + n_4 + n_5$, were investigated: 30, 60, and 120. Three levels of ratio of group sizes, ($n_1 / n_2 / n_3 / n_4 / n_5$: 1/1/1/1/1, 1/1/1/1/2 and 1/1/2/2/4) were used.

Population variance ratios

For the three group simulation, nine levels of variance ratios were investigated ($\sigma_1^2 / \sigma_2^2 / \sigma_3^2$: 1/1/4, 1/4/4, 1/1/2, 1/2/2, 1/1/1, 2/2/1, 2/1/1, 4/4/1, and 4/1/1). For the four group simulation, seven levels of variance ratios were investigated ($\sigma_1^2 / \sigma_2^2 / \sigma_3^2 / \sigma_4^2$: 1/1/4/4, 1/1/2/4, 1/1/1/2, 1/1/1/1, 2/1/1/1, 4/2/1/1, and 4/4/1/1). For the five group simulation, five levels of variance ratios were investigated (1/1/1/1/4, 1/1/1/1/2, 1/1/1/1/1, 2/1/1/1/1, and 4/1/1/1/1). Variance ratios were manipulated by multiplying the population of one or more of the groups in the design by a constant to create an imbalance in the variance ratios. The value of the constant was dependent on the amount of variance imbalance that was required for the cell of the design. For example, to create a variance ratio of 2/1/1, the scores of group whose variance is to be changed will have their variances adjusted by multiplying the selected group's variance by the square root of 2. The design was created so that there were direct pairing and inverse pairing in relation to unbalanced groups and direction of variance imbalance. Direct pairing occurs when the larger sample sizes are paired with the larger variance and inverse pairing occurs when the smaller sample size is paired with the larger variance (Tomarken & Serlin, 1986). This was done to investigate a more complete range of data possibilities. In addition, Keyes and Levy (1997) drew our attention to concern with unequal sample sizes, particularly in the case of factorial designs – see also O'Brien (1978, 1979) for discussion of Levene's test in additive models for variances. Findings suggest that the validity and efficiency of a statistical test is somewhat dependent on the direction of the pairing of sample sizes with the ratio of variance.

As a whole, the complex multivariate variable space represented by our simulation design captures many of the possibilities that might be found in day-to-day research practice.

Determining Type I Error Rates & Power

The frequency of Type I errors was tabulated for each cell in the design. For the three, four, and five group simulations, there were 540, 588, and 180 cells in each of the simulation designs respectively. As a description of our methodology, the following will describe the procedure for the ML and NPL tests for completing the steps for one cell in the design for the three group case as its description is generalizable to the four and five group scenarios. First, for both tests, three similarly distributed populations are generated and sampled from; for this example, it was three normally distributed populations that were sampled to create three groups. In this cell of the simulation design, each group had 10 members, and the population variances of the three groups are equal. This example tests the Type I errors for the two tests under the current conditions on the same set of data. For the ML, the absolute deviation from the median is calculated for each value in the sampled distribution and a one-way ANOVA is performed on these values to test if the variances are significantly different at the nominal alpha value of .05 (±.025). For the NPL, values are pooled and ranked, then partitioned back into their respective groups. A one-way ANOVA is then performed on the ranked data of the three groups to determine if the variances are statistically significantly different at the nominal alpha value of .05 (±.025). The value of ±.025 represents a liberal indicator of robustness and comes from Bradley (1978). The choice of Bradley's criterion is somewhat arbitrary, although it is the most liberal choice between the alternatives, and some of our conclusions may change with the other criteria. It should be noted that when Type I error rates are less than .05, the validity of the test is not jeopardized to the same extent as they are when they are inflated. This makes a test invalid if the rate of Type I errors are inflated, but when they decrease, the test becomes more conservative, reducing power. Reducing power does not invalidate the results of a test, so tests will be considered to be invalid only if the Type I error rate is inflated. This procedure was replicated 5000 times for each cell in the design.

In the cells where the ratio of variances was not equal and that maintained their Type I error rates, statistical power is represented by the proportion of times that the ML test, and the NPL test, correctly rejected the null hypothesis.

RESULTS

Three group simulation

The Type I error rates for the ML test and the NPL test for all of the conditions in the study are illustrated in Table 1. In all of the conditions of the simulation, both tests maintain their Type I error rate, with the ML test being somewhat conservative in many of the conditions. For example, the first row in Table 1 (reading across the row left to right), for a skew of 0, with an overall sample size of 30 with $n_1/n_2/n_3$ = (5/5/20), the Type I error rate for the NPL test is .056 and the Type I error rate for the ML test is .022.

Table 1. Three group Type I error rates of the Nonparametric and Median versions of the Levene tests under equivalent variance conditions.

		Skew = 0		Skew = 1		Skew = 2		Skew = 3	
N	n1/n2/n3	NPL	ML	NPL	ML	NPL	ML	NPL	ML
30	5/5/20	.056	.022	.057	.023	.060	.027	.059	.038
30	5/10/15	.057	.027	.056	.027	.058	.039	.059	.033
30	10/10/10	.047	.033	.050	.041	.050	.049	.048	.058
60	10/10/40	.054	.033	.058	.041	.057	.038	.058	.048
60	10/20/30	.047	.034	.056	.045	.055	.050	.048	.047
60	20/20/20	.047	.032	.051	.043	.049	.042	.045	.046
90	15/15/60	.050	.036	.052	.035	.062	.043	.056	.046
90	15/30/45	.050	.040	.048	.040	.051	.045	.051	.043
90	30/30/30	.055	.044	.053	.045	.050	.051	.053	.046

Note. NPL = Nonparametric Levene; ML = Median Levene.

It was the case that the Type I error rates of both tests was maintained in all of the conditions of the present study, thus power values for all of the simulated conditions will be reported. Table 2 reports the power values of the ML test and the NPL tests when the population skew is equal to 0. In nearly all of the cells of the Table 2 the two tests performed in a similar nature. For example, in the first row of the table are the results for the NPL test, which, for a sample size of 30 with $n_1/n_2/n_3$ = (5/5/20), and a ratio of variances of 1/1/4), the power is .385; that is, 38.5 percent of the null hypotheses were correctly rejected. In comparison, the power of the ML test (the next row in the table) under the same conditions was .247. When the total sample size was 30 the NPL test had a slight power advantage over the ML test in many of the cells of the design (i.e., 18 of the 24 cells in this section of the design); however, these power differences were small and in the cases when the ML had a power advantage, the differences were also small. When the sample size increased to 60 the power values of the two tests were very similar. When sample sizes were 90, the ML had a power advantage of the NPL in many of the cells of the design.

Table 2. Three group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of zero.

			Population Variance Ratio							
			Direct Pairings					Indirect Pairings		
Test	N	n1/n2/n3	1/1/4	1/4/4	1/1/2	1/2/2	2/2/1	2/1/1	4/4/1	4/1/1
NPL	30	5/5/20	.385	.218	.152	.111	.085	.082	.235	.158
ML	30	5/5/20	.247	.077	.069	.039	.055	.047	.260	.194
NPL	30	5/10/15	.377	.205	.148	.092	.106	.076	.302	.153
ML	30	5/10/15	.330	.085	.088	.040	.080	.053	.305	.197
NPL	30	10/10/10	.298	.293	.111	.094	.100	.107	.289	.284
ML	30	10/10/10	.381	.241	.099	.076	.081	.102	.241	.347
NPL	60	10/10/40	.705	.457	.255	.163	.154	.119	.485	.321
ML	60	10/10/40	.706	.343	.201	.111	.236	.174	.745	.594
NPL	60	10/20/30	.705	.456	.247	.147	.190	.111	.620	.321
ML	60	10/20/30	.796	.349	.263	.105	.240	.170	.775	.572
NPL	60	20/20/20	.581	.629	.193	.189	.193	.199	.624	.567
ML	60	20/20/20	.772	.673	.244	.189	.196	.249	.675	.760
NPL	90	15/15/60	.869	.676	.338	.221	.236	.166	.707	.495
ML	90	15/15/60	.920	.638	.334	.178	.345	.259	.909	.779
NPL	90	15/30/45	.882	.653	.339	.203	.293	.159	.836	.504
ML	90	15/30/45	.956	.622	.397	.170	.380	.251	.945	.780
NPL	90	30/30/30	.792	.834	.288	.291	.280	.267	.838	.787
ML	90	30/30/30	.932	.906	.383	.325	.319	.364	.915	.936

Note. NPL = Nonparametric Levene; ML = Median Levene.

The next condition investigated in the three group simulation was where the skew of the population distribution was equal to 1. Table 3 illustrates the power values of the NPL and the ML tests. When the sample size was 30, the NPL had small to moderate power differences with the ML test. For example, in Table 3 for the condition where N=30, n1/n2/n3 = 5/5/20 and the variance ratio is 1/1/4, the NPL has a power value of .424 and the ML has a power value of .184. In 23 of the 24 cells when the total sample size was equal to 30, the NPL possessed higher power values than the ML. As sample size increased to 60 and 90, the power differences between the two tests become smaller with the two tests performing quite similarly across the cells of the design.

Table 3. Three group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of one.

			Population Variance Ratio							
			Direct Pairings					Indirect Pairings		
Test	N	n1/n2/n3	1/1/4	1/4/4	1/1/2	1/2/2	2/2/1	2/1/1	4/4/1	4/1/1
NPL	30	5/5/20	.424	.249	.166	.111	.099	.087	.256	.226
ML	30	5/5/20	.184	.072	.055	.036	.061	.050	.177	.182
NPL	30	5/10/15	.425	.235	.157	.107	.126	.088	.338	.162
ML	30	5/10/15	.262	.081	.075	.044	.090	.058	.266	.171
NPL	30	10/10/10	.316	.327	.127	.120	.128	.122	.330	.333
ML	30	10/10/10	.313	.203	.090	.081	.085	.100	.206	.316
NPL	60	10/10/40	.765	.523	.298	.177	.175	.143	.545	.374
ML	60	10/10/40	.547	.252	.156	.092	.211	.160	.668	.517
NPL	60	10/20/30	.774	.516	.282	.173	.229	.128	.695	.360
ML	60	10/20/30	.691	.258	.207	.083	.209	.154	.677	.512
NPL	60	20/20/20	.642	.708	.238	.221	.226	.237	.705	.648
ML	60	20/20/20	.669	.563	.211	.155	.161	.201	.554	.681
NPL	90	15/15/60	.922	.750	.399	.252	.280	.182	.765	.542
ML	90	15/15/60	.821	.481	.259	.135	.301	.211	.851	.690
NPL	90	15/30/45	.929	.733	.419	.244	.353	.190	.893	.549
ML	90	15/30/45	.897	.469	.329	.133	.315	.210	.888	.701
NPL	90	30/30/30	.840	.885	.337	.335	.340	.342	.890	.846
ML	90	30/30/30	.870	.804	.321	.257	.262	.317	.803	.882

Note. NPL = Nonparametric Levene; ML = Median Levene.

The power values of the three group case where the skew of the population distribution is equal to 2 are listed in table 4. The NPL had higher power values than the ML in every cell of this part of the design. The magnitude of the power differences between the two tests ranged from moderate to large. For example, in the condition where N = 30, n1/n2/n3 was 5/5/20 and the variance ratio was 1/1/4, the power of the NPL was .556, whereas the power for the ML was .092.

Table 5 lists the power values of the ML and the NPL tests when the skew of the population distribution is equal to 3. The NPL possessed higher power values that the ML in every cell with power differences that are generally large. For example, in the condition where N = 30, n1/n2/n3 was 5/5/20 and the variance ratio was 1/1/4, the power of the NPL was .713, whereas the power for the ML was .025.

Four group simulation

The Type I error rates for the NPL and ML are presented in Table 6. Type I error rates were maintained in every cell in the four group simulation. It should be noted that the Type I error of the NPL exceed .07 in few of the cells, but stayed within the bounds of .075 allowing for the interpretation of the power values. For example, the condition where the total sample size is 40, n1/n2/n3/n4 4/4/16/16, the NPL has a Type I error rate of .070 and the ML has a Type I error rate of .061.

Table 4. Three group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of two.

			Population Variance Ratio							
			Direct Pairings				Indirect Pairings			
Test	N	n1/n2/n3	1/1/4	1/4/4	1/1/2	1/2/2	2/2/1	2/1/1	4/4/1	4/1/1
NPL	30	5/5/20	.556	.335	.238	.159	.141	.118	.333	.230
ML	30	5/5/20	.092	.053	.039	.035	.080	.060	.199	.146
NPL	30	5/10/15	.543	.314	.230	.132	.188	.108	.437	.211
ML	30	5/10/15	.155	.060	.054	.043	.079	.067	.198	.158
NPL	30	10/10/10	.415	.439	.180	.178	.175	.177	.442	.404
ML	30	10/10/10	.220	.159	.084	.079	.073	.079	.149	.211
NPL	60	10/10/40	.904	.694	.451	.276	.288	.197	.686	.441
ML	60	10/10/40	.295	.132	.080	.064	.185	.126	.510	.383
NPL	60	10/20/30	.893	.684	.454	.260	.376	.193	.841	.433
ML	60	10/20/30	.460	.127	.126	.064	.147	.135	.481	.376
NPL	60	20/20/20	.759	.860	.356	.364	.365	.361	.857	.768
ML	60	20/20/20	.464	.339	.145	.108	.116	.144	.334	.483
NPL	90	15/15/60	.985	.902	.618	.405	.454	.308	.875	.659
ML	90	15/15/60	.532	.247	.137	.090	.228	.168	.691	.517
NPL	90	15/30/45	.981	.890	.642	.392	.566	.288	.968	.660
ML	90	15/30/45	.679	.246	.193	.084	.217	.170	.685	.535
NPL	90	30/30/30	.929	.971	.526	.567	.547	.514	.974	.927
ML	90	30/30/30	.694	.561	.201	.170	.159	.208	.565	.680

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 7 presents the power values of the two tests when the skew of the population distribution is equal to 0. Overall, power differences between the NPL and the ML are small. The NPL has a small power advantage over the ML in 16 of the 24 cells in the condition where total sample size is equal to 40. For example, in the condition where the total sample size is 40, n1/n2/n3/n4 is 4/4/16/16 and the ratio of variances is 1/1/2/4, the power of the NPL is .285 and the power for the ML is .189. When the total sample size is equal to 80 or 120, the ML has a small to moderate power advantage over the ML in 45 of the 48 cells. For example, when the total sample size is 80, n1/n2/n3/n/4 is 8/8/32/32 and the ratio of variances is 4/4/1/1 the NPL has a power value of .374, whereas the ML's power is equal to .692.

Table 5. Three group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of three.

			Population Variance Ratio							
			Direct Pairings				Indirect Pairings			
Test	N	n1/n2/n3	1/1/4	1/4/4	1/1/2	1/2/2	2/2/1	2/1/1	4/4/1	4/1/1
NPL	30	5/5/20	.713	.463	.510	.303	.307	.197	.483	.336
ML	30	5/5/20	.025	.036	.021	.029	.089	.062	.167	.125
NPL	30	5/10/15	.674	.450	.498	.273	.399	.201	.621	.305
ML	30	5/10/15	.062	.040	.036	.040	.072	.069	.137	.123
NPL	30	10/10/10	.531	.629	.365	.397	.395	.359	.626	.540
ML	30	10/10/10	.127	.097	.071	.064	.064	.069	.096	.120
NPL	60	10/10/40	.989	.914	.866	.619	.616	.388	.837	.563
ML	60	10/10/40	.065	.060	.029	.038	.139	.096	.332	.242
NPL	60	10/20/30	.964	.908	.852	.613	.776	.395	.948	.560
ML	60	10/20/30	.165	.055	.053	.040	.106	.105	.274	.242
NPL	60	20/20/20	.867	.978	.717	.795	.801	.716	.975	.865
ML	60	20/20/20	.231	.161	.098	.072	.073	.087	.173	.231
NPL	90	15/15/60	1.000	.991	.966	.838	.835	.589	.965	.766
ML	90	15/15/60	.157	.079	.043	.048	.167	.120	.462	.328
NPL	90	15/30/45	.998	.991	.968	.833	.944	.581	.997	.757
ML	90	15/30/45	.304	.088	.088	.050	.146	.127	.384	.334
NPL	90	30/30/30	.967	.999	.895	.947	.947	.901	.999	.966
ML	90	30/30/30	.376	.253	.120	.098	.097	.119	.260	.355

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 6. Four group Type I error rates of the Nonparametric and Median versions of the Levene tests under equivalent variance conditions.

		Skew = 0		Skew = 1		Skew = 2		Skew = 3	
N	n1/n2/n3/n4	NPL	ML	NPL	ML	NPL	ML	NPL	ML
40	4/4/16/16	.070	.039	.071	.049	.069	.055	.066	.050
40	5/5/10/20	.061	.019	.064	.023	.057	.029	.056	.042
40	8/8/8/16	.060	.030	.055	.040	.056	.052	.056	.054
40	10/10/10/10	.053	.033	.058	.043	.055	.046	.050	.052
80	8/8/32/32	.054	.034	.060	.038	.064	.044	.060	.049
80	10/10/20/40	.050	.038	.053	.041	.061	.045	.051	.045
80	16/16/16/32	.046	.031	.056	.039	.054	.051	.053	.045
80	20/20/20/20	.052	.030	.051	.037	.058	.042	.046	.048
120	12/12/48/48	.055	.041	.056	.041	.059	.047	.055	.049
120	15/15/30/60	.053	.038	.059	.040	.057	.046	.052	.050
120	24/24/24/48	.051	.040	.051	.045	.048	.046	.052	.043
120	30/30/30/30	.057	.045	.045	.043	.051	.045	.054	.046

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 8 lists the power values of the NPL and the ML tests when the skew of the population distribution is equal to 1. The NPL has a small to moderate power advantage over the ML in 20 of the 24 cells when the sample size is equal to 40. For example, when the total sample size is 40, n1/n2/n3/n4 is 4/4/16/16 and the ratio of variances is 1/1/2/4, the NPL has a power value of .332, whereas the ML has a power value of .148.

Table 7. Four group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of zero.

			Population Variance Ratio					
			Direct Parings			Indirect Pairings		
Test	N	n1/n2/n3/n4	1/1/4/4	1/1/2/4	1/1/1/2	2/1/1/1	4/2/1/1	4/4/1/1
NPL	40	4/4/16/16	.303	.285	.159	.073	.145	.180
ML	40	4/4/16/16	.114	.189	.118	.078	.265	.381
NPL	40	5/5/10/20	.306	.299	.141	.081	.150	.218
ML	40	5/5/10/20	.142	.180	.066	.036	.144	.210
NPL	40	8/8/8/16	.397	.356	.151	.091	.226	.328
ML	40	8/8/8/16	.338	.322	.105	.090	.304	.419
NPL	40	10/10/10/10	.375	.271	.113	.107	.258	.371
ML	40	10/10/10/10	.394	.321	.105	.097	.302	.388
NPL	80	8/8/32/32	.626	.543	.269	.100	.256	.374
ML	80	8/8/32/32	.510	.568	.289	.140	.509	.692
NPL	80	10/10/20/40	.638	.569	.240	.103	.310	.455
ML	80	10/10/20/40	.594	.599	.245	.144	.539	.723
NPL	80	16/16/16/32	.742	.672	.258	.152	.481	.672
ML	80	16/16/16/32	.812	.774	.283	.209	.691	.857
NPL	80	20/20/20/20	.734	.555	.180	.188	.539	.738
ML	80	20/20/20/20	.860	.716	.228	.252	.711	.856
NPL	120	12/12/48/48	.821	.739	.386	.109	.398	.577
ML	120	12/12/48/48	.811	.817	.473	.188	.722	.887
NPL	120	15/15/30/60	.875	.834	.403	.143	.495	.706
ML	120	15/15/30/60	.903	.896	.465	.223	.785	.931
NPL	120	24/24/24/48	.919	.870	.363	.235	.700	.874
ML	120	24/24/24/48	.967	.952	.456	.349	.893	.976
NPL	120	30/30/30/30	.915	.768	.266	.283	.765	.920
ML	120	30/30/30/30	.978	.914	.373	.391	.913	.982

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 9 presents the power values of the two tests when the skew of the population distribution is equal to 2. The NPL has moderate to large power advantages over the ML in nearly every cell of the design. For example, when the total sample size is 40, n1/n2/n3/n4 is 4/4/16/16 and the ratio of variances is 1/1/2/4, the power of the NPL is .464 and the power of the ML is .094. In the conditions where the total sample size is 80 and 120, the NPL has small to moderate power differentials with the ML. For example, when the total sample size is 80, n1/n2/n3/n4 is 10/10/20/40 and the ratio of variance is 4/4/1/1, the NPL's power is .651, whereas the ML's power is .484.

Table 8. Four group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of one.

			Population Variance Ratio					
			Direct Parings			Indirect Pairings		
Test	N	n1/n2/n3/n4	1/1/4/4	1/1/2/4	1/1/1/2	2/1/1/1	4/2/1/1	4/4/1/1
NPL	40	4/4/16/16	.345	.332	.187	.087	.156	.206
ML	40	4/4/16/16	.096	.148	.100	.094	.243	.340
NPL	40	5/5/10/20	.347	.339	.168	.086	.185	.244
ML	40	5/5/10/20	.118	.140	.064	.045	.142	.186
NPL	40	8/8/8/16	.441	.401	.175	.105	.269	.362
ML	40	8/8/8/16	.270	.262	.103	.091	.276	.359
NPL	40	10/10/10/10	.419	.310	.127	.126	.300	.421
ML	40	10/10/10/10	.324	.256	.099	.096	.258	.325
NPL	80	8/8/32/32	.695	.619	.320	.099	.285	.426
ML	80	8/8/32/32	.355	.437	.223	.126	.439	.622
NPL	80	10/10/20/40	.701	.634	.276	.128	.351	.513
ML	80	10/10/20/40	.438	.460	.185	.138	.447	.627
NPL	80	16/16/16/32	.818	.747	.311	.180	.539	.742
ML	80	16/16/16/32	.687	.643	.229	.177	.596	.765
NPL	80	20/20/20/20	.790	.616	.221	.211	.631	.800
ML	80	20/20/20/20	.747	.601	.206	.193	.617	.751
NPL	120	12/12/48/48	.887	.830	.449	.142	.458	.636
ML	120	12/12/48/48	.650	.693	.367	.184	.633	.808
NPL	120	15/15/30/60	.924	.900	.487	.182	.571	.772
ML	120	15/15/30/60	.768	.790	.360	.199	.705	.867
NPL	120	24/24/24/48	.962	.915	.435	.270	.771	.912
ML	120	24/24/24/48	.915	.871	.364	.274	.809	.931
NPL	120	30/30/30/30	.952	.829	.319	.340	.827	.949
ML	120	30/30/30/30	.937	.832	.303	.310	.832	.933

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 10 lists the power values of the two tests when the skew of the population distribution is equal to 3. The NPL possesses moderate to large power advantages over the ML. For example, in the condition where the total sample size is 40, n1/n2/n3/n4 is 8/8/8/16 and the ratio of variances is 4/4/1/1, the NPL has a power value of .656 and the ML's power is equal to .173. When the total sample size is 80 or 120, the NPL is more powerful than the ML in every cell of the design and in many cases the power difference is very large. For example, when the total sample size is 80, n1/n2/n3/n4 is 8/8/32/32 and the ratio of variances is 4/4/1/1, the power of the NPL is .697 and the power of the ML is .335.

Table 9. Four group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of two.

			Population Variance Ratio					
			Direct Parings			Indirect Pairings		
Test	N	n1/n2/n3/n4	1/1/4/4	1/1/2/4	1/1/1/2	2/1/1/1	4/2/1/1	4/4/1/1
NPL	40	4/4/16/16	.461	.464	.282	.101	.200	.266
ML	40	4/4/16/16	.056	.094	.080	.091	.214	.289
NPL	40	5/5/10/20	.477	.453	.241	.108	.240	.324
ML	40	5/5/10/20	.080	.089	.047	.047	.128	.173
NPL	40	8/8/8/16	.589	.548	.257	.143	.359	.470
ML	40	8/8/8/16	.161	.167	.072	.081	.209	.260
NPL	40	10/10/10/10	.554	.426	.173	.183	.419	.546
ML	40	10/10/10/10	.226	.189	.085	.089	.173	.233
NPL	80	8/8/32/32	.861	.814	.499	.153	.388	.544
ML	80	8/8/32/32	.162	.236	.141	.109	.341	.485
NPL	80	10/10/20/40	.870	.831	.452	.175	.497	.651
ML	80	10/10/20/40	.239	.263	.119	.110	.354	.484
NPL	80	16/16/16/32	.934	.896	.476	.295	.712	.863
ML	80	16/16/16/32	.426	.397	.140	.142	.423	.545
NPL	80	20/20/20/20	.921	.791	.337	.336	.792	.911
ML	80	20/20/20/20	.521	.414	.144	.142	.413	.527
NPL	120	12/12/48/48	.979	.959	.682	.213	.590	.768
ML	120	12/12/48/48	.330	.424	.228	.150	.476	.663
NPL	120	15/15/30/60	.987	.980	.724	.257	.718	.860
ML	120	15/15/30/60	.463	.494	.199	.162	.519	.712
NPL	120	24/24/24/48	.992	.987	.670	.421	.898	.981
ML	120	24/24/24/48	.699	.644	.218	.194	.609	.781
NPL	120	30/30/30/30	.991	.944	.510	.502	.946	.990
ML	120	30/30/30/30	.763	.620	.215	.207	.625	.768

Note. NPL = Nonparametric Levene; ML = Median Levene.

Five group simulation

Table 11 lists the Type I error rates for the ML and the NPL tests. Once again, the nominal Type I error rate was maintained for both tests in every cell of the design. The NPL did have some slightly elevated error rates in some of the cells of the design compared to the ML; however, these values are within the liberal criteria for robustness. For example, when the total sample size is 30, n1/n2/n3/n4/n5 is 3/3/6/6/12 and skew is zero, the NPL has a Type I error rate of .071 and the ML has a Type I error rate of .021.

Table 12 presents the power values for the ML and the NPL tests when the skew of the population distribution is equal to 0. When the sample size was small the NPL has a small to moderate power advantage of the ML. For example, when the total sample size is 30, n1/n2/n3/n4/n5 is 3/3/6/6/12 and the ratio of variances is 1/1/1/1/4, the NPL has a power value of .317 and the ML's power is.189. When the overall sample size 60 or 90, the ML possesses small to moderate power advantage over the NPL in most cells. For example, when the overall sample size is 60, n1/n2/n3/n4/n5 is 6/6/12/12/24 and the variance ratio is 1/1/1/1/4, the power of the NPL is .576 and the power of the ML is .63. Overall, when the skew was equal to zero both tests performed similarly with the NPL performing slightly better when the sample sizes was small and the ML performing better when the sample sizes were larger.

Table 10. Four group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of three.

			Population Variance Ratio					
			Direct Parings			Indirect Pairings		
Test	N	n1/n2/n3/n4	1/1/4/4	1/1/2/4	1/1/1/2	2/1/1/1	4/2/1/1	4/4/1/1
NPL	40	4/4/16/16	.644	.757	.558	.162	.345	.407
ML	40	4/4/16/16	.029	.046	.059	.094	.189	.242
NPL	40	5/5/10/20	.640	.694	.502	.192	.393	.453
ML	40	5/5/10/20	.045	.035	.044	.062	.112	.142
NPL	40	8/8/8/16	.776	.756	.542	.267	.572	.656
ML	40	8/8/8/16	.080	.073	.047	.074	.150	.173
NPL	40	10/10/10/10	.734	.664	.349	.337	.643	.721
ML	40	10/10/10/10	.127	.121	.063	.068	.108	.120
NPL	80	8/8/32/32	.981	.982	.882	.286	.605	.697
ML	80	8/8/32/32	.045	.081	.069	.100	.250	.335
NPL	80	10/10/20/40	.982	.971	.836	.356	.706	.809
ML	80	10/10/20/40	.079	.090	.048	.087	.219	.305
NPL	80	16/16/16/32	.992	.984	.885	.576	.904	.963
ML	80	16/16/16/32	.175	.160	.069	.094	.242	.301
NPL	80	20/20/20/20	.984	.954	.681	.689	.951	.982
ML	80	20/20/20/20	.253	.207	.091	.092	.197	.248
NPL	120	12/12/48/48	1.000	1.000	.976	.426	.814	.888
ML	120	12/12/48/48	.083	.147	.104	.114	.330	.441
NPL	120	15/15/30/60	1.000	1.000	.989	.549	.897	.950
ML	120	15/15/30/60	.136	.161	.075	.118	.337	.452
NPL	120	24/24/24/48	1.000	1.000	.978	.788	.984	.995
ML	120	24/24/24/48	.306	.282	.096	.117	.352	.458
NPL	120	30/30/30/30	.999	.996	.876	.875	.994	1.000
ML	120	30/30/30/30	.404	.311	.118	.118	.316	.407

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 13 lists the power values for the two tests when the skew of the population distribution is equal to 1. The NPL had a power advantage over the ML in every cell in the table except for two. The differences in power between the two tests are small in some cases (e.g., when N = 30, n1/n2/n3/n4/n5 = 3/3/6/6/12 and the variance ratio is 1/1/1/1/2, the NPL has a power value of .112 whereas the ML has a power value of .047. In many cells the differences in power are quite large (e.g., when N = 30, n1/n2/n3/n4/n5 = 6/6/6/6/6 and the variance ratio is 1/1/1/1/2, the NPL has a power value of .527 and the ML's power is .072.

Table 11. Five group Type I error rates of the Nonparametric and Median versions of the Levene tests under equivalent variance conditions.

		Skew = 0		Skew = 1		Skew = 2		Skew = 3	
N	n1/n2/n3/n4/n5	NPL	ML	NPL	ML	NPL	ML	NPL	ML
30	3/3/6/6/12	.071	.021	.069	.023	.071	.032	.070	.041
30	5/5/5/5/10	.071	.006	.068	.010	.073	.021	.075	.033
30	6/6/6/6/6	.065	.035	.065	.047	.066	.069	.068	.069
60	6/6/12/12/24	.064	.033	.056	.034	.058	.045	.064	.055
60	10/10/10/10/20	.051	.028	.054	.037	.059	.046	.063	.051
60	12/12/12/12/12	.061	.027	.053	.035	.059	.047	.060	.049
90	9/9/18/18/36	.062	.033	.060	.035	.060	.042	.058	.046
90	15/15/15/15/30	.052	.027	.057	.035	.055	.042	.055	.045
90	18/18/18/18/18	.052	.035	.051	.038	.051	.043	.050	.042

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 12. Five group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of zero.

			Population Variance Ratio			
			Direct Pairings		Indirect Parings	
Test	N	n1/n2/n3/n4/n5	1/1/1/1/4	1/1/1/1/2	2/1/1/1/1	4/1/1/1/1
NPL	30	3/3/6/6/12	.317	.140	.076	.096
ML	30	3/3/6/6/12	.189	.047	.016	.042
NPL	30	5/5/5/5/10	.277	.129	.088	.161
ML	30	5/5/5/5/10	.181	.041	.015	.091
NPL	30	6/6/6/6/6	.166	.094	.094	.175
ML	30	6/6/6/6/6	.198	.068	.073	.198
NPL	60	6/6/12/12/24	.576	.215	.080	.172
ML	60	6/6/12/12/24	.630	.170	.089	.304
NPL	60	10/10/10/10/20	.516	.178	.103	.277
ML	60	10/10/10/10/20	.617	.154	.109	.460
NPL	60	12/12/12/12/12	.321	.119	.117	.335
ML	60	12/12/12/12/12	.489	.118	.111	.512
NPL	90	9/9/18/18/36	.788	.278	.098	.238
ML	90	9/9/18/18/36	.889	.281	.114	.466
NPL	90	15/15/15/15/30	.729	.241	.143	.415
ML	90	15/15/15/15/30	.866	.258	.174	.673
NPL	90	18/18/18/18/18	.496	.162	.169	.484
ML	90	18/18/18/18/18	.737	.202	.208	.736

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 13. Five group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of one.

			Population Variance Ratio			
			Direct Pairings		Indirect Parings	
Test	N	n1/n2/n3/n4/n5	1/1/1/1/4	1/1/1/1/2	2/1/1/1/1	4/1/1/1/1
NPL	30	3/3/6/6/12	.319	.112	.333	.323
ML	30	3/3/6/6/12	.191	.047	.028	.045
NPL	30	5/5/5/5/10	.300	.297	.520	.536
ML	30	5/5/5/5/10	.149	.043	.020	.080
NPL	30	6/6/6/6/6	.537	.527	.534	.535
ML	30	6/6/6/6/6	.181	.072	.060	.183
NPL	60	6/6/12/12/24	.588	.285	.991	.990
ML	60	6/6/12/12/24	.636	.126	.090	.278
NPL	60	10/10/10/10/20	.884	.885	.999	.999
ML	60	10/10/10/10/20	.501	.116	.102	.373
NPL	60	12/12/12/12/12	1.000	1.000	1.000	.999
ML	60	12/12/12/12/12	.411	.113	.109	.426
NPL	90	9/9/18/18/36	.779	.600	1.000	1.000
ML	90	9/9/18/18/36	.884	.212	.106	.411
NPL	90	15/15/15/15/30	.999	.999	1.000	1.000
ML	90	15/15/15/15/30	.763	.199	.144	.565
NPL	90	18/18/18/18/18	1.000	1.000	1.000	1.000
ML	90	18/18/18/18/18	.632	.165	.167	.645

Note. NPL = Nonparametric Levene; ML = Median Levene.

Table 14 lists the power values for the two tests when the skew of the population distribution is equal to 2. The NPL demonstrated a power advantage over the ML in every cell of the table. Once again the power differences ranged from small to large. For example, when the total sample size was 30, the ratio of sample sizes was 3/3/6/6/12 and the ratio of variances was 1/1/1/1/2, the NPL had a small power advantage over the ML with values of .113 and .044 respectively. Whereas, when the total sample size was 30, the ratio of sample sizes was 3/3/6/6/12 and the ratio of variances was 1/1/1/1/2, the NPL had quite a large power advantage over the ML with values of .529 and .080 respectively.

Table 15 lists the power values for the two tests when the skew of the population distribution is equal to 3. Once again, the NPL possessed a power advantage over the ML in every cell of the table. Once again the power differences ranged from small to large. For example, when the total sample size was 30, the ratio of sample sizes was 3/3/6/6/12 and the ratio of variances was 1/1/1/1/2, the NPL had a small power advantage over the ML with values of .119 and .039 respectively. Whereas, when the total sample size was 30, the ratio of sample sizes was 3/3/6/6/12 and the ratio of variances was 1/1/1/1/2, the NPL had quite a large power advantage over the ML with values of .534 and .081 respectively.

Table 14. Five group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of two.

			Population Variance Ratio			
			Direct Pairings		Indirect Parings	
Test	N	n1/n2/n3/n4/n5	1/1/1/1/4	1/1/1/1/2	2/1/1/1/1	4/1/1/1/1
NPL	30	3/3/6/6/12	.117	.113	.324	.329
ML	30	3/3/6/6/12	.092	.044	.033	.057
NPL	30	5/5/5/5/10	.298	.296	.525	.540
ML	30	5/5/5/5/10	.107	.038	.031	.068
NPL	30	6/6/6/6/6	.524	.529	.519	.529
ML	30	6/6/6/6/6	.141	.080	.073	.151
NPL	60	6/6/12/12/24	.274	.282	.990	.991
ML	60	6/6/12/12/24	.288	.074	.087	.227
NPL	60	10/10/10/10/20	.883	.885	1.000	1.000
ML	60	10/10/10/10/20	.316	.094	.085	.266
NPL	60	12/12/12/12/12	1.000	1.000	1.000	1.000
ML	60	12/12/12/12/12	.286	.091	.091	.290
NPL	90	9/9/18/18/36	.621	.608	1.000	1.000
ML	90	9/9/18/18/36	.510	.126	.100	.288
NPL	90	15/15/15/15/30	.998	.999	1.000	1.000
ML	90	15/15/15/15/30	.508	.129	.107	.407
NPL	90	18/18/18/18/18	1.000	1.000	1.000	1.000
ML	90	18/18/18/18/18	.457	.124	.115	.453

Note. NPL = Nonparametric Levene; ML = Median Levene.

DISCUSSION

The findings from the series of simulations that were conducted provide further support for the usefulness of the NPL when data are sampled from distributions that tend to be more heavily skewed. In general, the Type I error rates of the ML tended to be consistently lower than the NPL; however, the overly conservative nature of the ML tends to result in lower power values, which was demonstrated in the current simulations. In some of the cells in the current simulation design, the NPL had slightly elevated Type I error rates in comparison to the ML; however, they remained within the liberal criteria set out by Bradley (1978). Results support the utility of the NPL across a wide variety of ANOVA designs, especially when sample sizes are small and population distributions may be skewed or unknown.

Table 15. Five group power values of the Nonparametric and Median version of the Levene test for equality of variances for skew of three.

			Population Variance Ratio			
			Direct Pairings		Indirect Parings	
Test	N	n1/n2/n3/n4/n5	1/1/1/1/4	1/1/1/1/2	2/1/1/1/1	4/1/1/1/1
NPL	30	3/3/6/6/12	.114	.119	.330	.345
ML	30	3/3/6/6/12	.046	.039	.052	.065
NPL	30	5/5/5/5/10	.303	.297	.539	.528
ML	30	5/5/5/5/10	.059	.037	.039	.065
NPL	30	6/6/6/6/6	.526	.534	.524	.531
ML	30	6/6/6/6/6	.109	.081	.080	.112
NPL	60	6/6/12/12/24	.280	.275	.990	.993
ML	60	6/6/12/12/24	.091	.044	.082	.157
NPL	60	10/10/10/10/20	.895	.889	1.000	.999
ML	60	10/10/10/10/20	.112	.051	.078	.158
NPL	60	12/12/12/12/12	.999	.999	.999	.999
ML	60	12/12/12/12/12	.146	.070	.069	.164
NPL	90	9/9/18/18/36	.600	.619	1.000	1.000
ML	90	9/9/18/18/36	.177	.050	.086	.191
NPL	90	15/15/15/15/30	.998	.998	1.000	1.000
ML	90	15/15/15/15/30	.194	.064	.079	.231
NPL	90	18/18/18/18/18	1.000	1.000	1.000	1.000
ML	90	18/18/18/18/18	.233	.076	.082	.226

Note. NPL = Nonparametric Levene; ML = Median Levene.

When the overall sample sizes were in the larger two categories (e.g., 60 and 90 for the five group simulation) and the skew of the population distribution was equal to 0, the ML had an overall power advantage over the NPL; however, when the overall sample sizes were smaller and the skew of the distribution was 1 or larger, the NPL was consistently more powerful than the ML.

Interestingly, the power of both of the tests were impacted by the imbalance between the numbers in each group with more group imbalance leading to both increases and decreases in power. One pattern that tended to emerge in the results was that in the direct pairing condition, as the groups became more unbalanced the power of the NPL tended to increase and the power of the ML tended to decrease. Whereas, in the indirect pairing conditions, as the groups become imbalanced, the power of the ML tended to go up and the power of the NPL tended to decrease. This pattern was not consistent across all conditions but did tend to coincide with the conditions where skew was higher (i.e., 2 or 3). In addition, the magnitude of the differences in the variances between the groups impacted the results. This finding makes intuitive sense as the magnitude difference between the variances essentially represents the effect size for this simulation study.

More interesting is the interaction of ratio of sample sizes and the ratio of variances. Note that in terms of impact of direct versus indirect pairing between the degree of imbalance between the groups sizes and the degree of inequality of the variances, the findings support those of Nordstokke and Zumbo (2010) whereby the NPL had a power advantage when the pairings were direct and the ML had a power advantage when the pairings were indirect. As noted by Nordstokke and Zumbo (2010), the direction of pairing impacts the mean square values in the model resulting in distorted expressions of variance.

Even though the median version of the Levene test has been demonstrated to have good statistical properties and robustness, using it as the only comparison test reduces the generalizability of the results; however, future studies will include a broader spectrum of tests of variance (e.g., bootstrapping approaches) to further support the potential utility of the NPL. Nevertheless, the results of the current study are an important first step in establishing the usefulness of the NPL as a practical statistical tool that may be utilize in a wide variety of research settings where small sample sizes or skewed data are often found.

One caveat that was present in Nordstokke and Zumbo (2010) is also present in this paper relates to the generalizability of the results. Since only Chi-square distributions were used in the simulation study, the results could reflect some idiosyncrasy present within the data generation method. This was done purposefully to replicate the method used by Nordstokke and Zumbo (2010). As mentioned in that paper, this does not invalidate the findings of the present study, but instead illustrates that a wider variety of distributions need to be used in future studies. It is also important to note that this study used more liberal alpha criterion for assessing robustness than was used by Nordstokke and Zumbo (2010). In their study, .05 (± .01) and the current study used .05 (± .025). This allowed for a broader discussion of the results in terms of power; however, if the more strict criterion of .05 (± .01) had been used then there would have been several cells of the design where the Type I error of the NPL would have been elevated beyond the .06 level. This was evident in the four and five group cases and for the small sample size condition (e.g., N=40). A problem inherent the interpretation of simulation research of this nature is that no studies have been conducted that inform us on the limits of the allowable differences in variances for analysis of variance type tests. Put another way, we do not know what degree of variance heterogeneity (in combination with distributional disturbances, sample size, direct or inverse pairing of group size, etc.) is necessary to increase the Type I error rate of, for example, the ANOVA test of means to an unacceptable level. Future research will investigate these bounds so that less arbitrary criterion for simulation studies can be established.

A point that deserves attention at this juncture has to do with the precision of the results. This simulation study was based on 5000 replications and is intended to be used to inform about the statistical properties of the tests being investigated. The results that are presented are essentially point estimates of the "true" Type I errors and power of the tests under investigation and by are not presented as proof of the validity of the robustness of the NPL, but as evidence of its potential utility as a data analytic tool. Future studies will focus on investigating its further utility.

To summarize, the simulation results demonstrate the potential utility of the NPL when data come from heavily skewed population distributions. This supports the findings of Nordstokke and Zumbo (2010) where the NPL maintained its Type I error rate and possessed high power values when population distributions were heavily skewed. It is important to note that the NPL has higher power when the total sample size is small across the three simulation studies. This suggests that the NPL has utility for research settings that tend to have yield smaller sample sizes and group membership often tends to be imbalanced or when data tend to be heavily skewed. This often occurs in psychological and health based research setting where access to participant populations can be challenging due to small populations or limited access to participants from their populations of interest.

REFERENCES

Box, G. E. P. (1953). Non-normality and tests on variance. *Biometrika,* 40, 318–335.

Bradley, J.V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144-152.

Brown, M.B., & Forsythe, A.B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(*2*), 364-367.

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

Conover, W.J., Johnson, M.E., & Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, *23*(*4*), 351- 361.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32*, 675–701.

Keyes, T. M., & Levy, M. S. (1997). Analysis of Levene's test under design imbalance. *Journal of Educational and Behavioral Statistics, 22,* 227-236.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin et al. (Eds.), *Contributions to probability and statistics: Essay in honor of Harold Hotelling* (pp. 2 78-292). Stanford, CA: Stanford University Press.

Nordstokke, D.W., & Zumbo, B.D. (2007). A cautionary tale about Levene's tests for equality of variances. *Journal of Educational Research and Policy Studies*, *7*, 1-14.

Nordstokke, D.W., & Zumbo, B.D. (2010). A new non-parametric test for equal variances. *Psicologica*, *31*, 401-430.

Nordstokke, D.W., Zumbo, D.W., Cairns, S.L., & Saklofske, D.H. (2011). The operating characteristics of the nonparametric Levene test for equal variances with assessment and evaluation data. *Practical Assessment, Research, and Evaluation*, *16*(*5*), 1-8.

O'Brien, R. G. (1978). Robust Techniques for Testing Heterogeneity of Variance Effects in Factorial Designs. *Psychometrika, 43*, 327-344.

O'Brien, R. G. (1979). A General ANOVA Method for Robust Tests of Additive Models for Variances. *Journal of the American Statistical Association, 74*, 877-880.

Tomarken, A.J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin, 99*(*1*), 90-99.

Zimmerman, D.W. (1987). Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances. *Journal of Experimental Education*, *55*, 171-174.

Zimmerman, D.W. (2004). A note on preliminary test of equality of variances. *British Journal of Mathematical and Statistical Psychology*, *57*, 173-181.

Zimmerman, D.W. (2012). A note on consistency of non-parametric rank tests and related rank transformations. *British Journal of Mathematical and Statistical Psychology*, *65*, 122-144.

APPENDIX 1

SPSS syntax used to run the NPL for the present simulation study

```
*Creating Absolute Rank Difference Value for ANOVA/Nonparametric Levene.

GET FILE = 'G:\Input Files\Simulated_Population_x1x2x3.sav'.

SORT CASES  BY Ndraw.
SPLIT FILE LAYERED BY Ndraw.
RANK VARIABLES=dv (A)
 /RANK
 /PRINT=YES
 /TIES=MEAN.
SPLIT FILE OFF.

AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES OVERWRITEVARS=YES
 /BREAK=Ndraw group
 /Rdv_mean=MEAN(Rdv).

COMPUTE Rdifference=ABS(Rdv-Rdv_mean).
EXECUTE.

SAVE OUTFILE='G:\Input Files\Simulated_Population_x1x2x3.sav'
/KEEP=all /COMPRESSED.
EXECUTE.

*Running ANOVA for Nonparametric Levene.

GET FILE = 'G:\Input Files\Simulated_Population_x1x2x3.sav'.

SORT CASES BY Ndraw (A) .
SPLIT FILE by ndraw.
EXECUTE.

OMS
 /SELECT TABLES
/IF COMMANDS=['Oneway']
    SUBTYPES=['ANOVA']
 /DESTINATION FORMAT=SAV
             OUTFILE='G:\Input Files\Nonparametric_Results.sav'
             VIEWER=no.
ONEWAY
 Rdifference BY group
 /STATISTICS DESCRIPTIVES EFFECTS
 /MISSING ANALYSIS .

OMSEND.
EXECUTE.
```

(Manuscript received: 10 March 2014; accepted: 13 May 2014)