Permutation Tests in the Educational and Behavioral Sciences A Systematic Review

Article *in* Methodology European Journal of Research Methods for the Behavioral and Social Sciences · September 2014

DOI: 10.1027/1614-2241/a000067

4 authors:

Ming Huo
Northeast Normal University

Mieke Heyvaert
Fonds Wetenschappelijk Onderzoek; KU Leuven

Wim Van den Noortgate
KU Leuven

Patrick Onghena
KU Leuven

Please cite this article as:

Huo, M., Heyvaert, M., Van Den Noortgate, W., & Onghena, P. (2014). Permutation tests in the educational and behavioral sciences: A systematic review. *Methodology - European Journal of Research Methods for the Behavioral and Social Sciences, 10*, 43-59. doi:10.1027/1614-2241/a000067

Permutation tests in the educational and behavioral sciences: A systematic review

Huo, M. [1], Heyvaert, M. [1, 2], Van Den Noortgate, W. [1], & Onghena, P. [1]

[1] Faculty of Psychology and Educational Sciences - KU Leuven

[2] Postdoctoral Fellow of the Research Foundation - Flanders (Belgium)

Permutation tests in the educational and behavioral sciences: A systematic review

In many educational and behavioral studies, the assumptions of common parametric hypothesis tests (e.g., random samples from defined populations, sampled populations are normally distributed, and of equal variance) are considered implausible (e.g., Adams & Anthony, 1996; Edgington & Onghena, 2007; Hunter & May, 2003; Ludbrook, 1994; Manly, 1997). An alternative to using parametric tests is to use nonparametric rank tests (e.g., Wilcoxon-Mann-Whitney tests and Wilcoxon signed test). These tests transform the original data into ranks and obtain the probability distribution for a given test statistic under the null hypothesis (Lehmann & D'Abrera, 2006). However, nonparametric rank methods are not unproblematic either. A common criticism is that the use of nonparametric methods results in a loss of information when reducing the data to ranks, eventually resulting in a reduction of the statistical power and efficiency (Edgington & Onghena, 2007).

Because of the problems associated with parametric and standard nonparametric tests, permutation tests (PTs) have become popular alternatives (Edgington & Onghena, 2007; Good, 2005; Manly, 1997). PTs constitute a set of distribution-free statistical tests that calculate the probability of getting a value as extreme or more extreme than an obtained value of a test statistic under a null hypothesis by recalculating the test statistic for all or many rearrangements or divisions of the data. When all possible permutations of the data are exhaustively listed, the relevant PTs are often called *systematic* PTs. When the total number of permutations is too large, *random (Monte Carlo)* PTs are usually carried out. These tests use only a subset of all possible data permutations to approximate the *p* value (Dwass, 1957; Edgington & Onghena, 2007; Manly, 1997). Although some researchers (e.g., Edgington & Onghena, 2007; Kempthorne & Doerfler, 1969) suggested distinguishing tests on designs with random assignment from those without random assignment by calling them "randomization tests" and "permutation tests", respectively, we choose not to make this distinction in this review. PTs do not depend on a specific error distribution, and they use the original values of the data instead of the ranks. PTs were proposed in the early twentieth century, but were not widely used until much later. This is mostly because (a) they were too computationally intensive, (b) their applicability was limited to simple scenarios, (c) and they could be replaced by the available classical nonparametric tests based on ranks (Welch, 1990).

Now PTs are experiencing a renaissance with the emergence of cheap and fast computers and applicable software, and with methodological and statistical advancements, making them applicable in a variety of experimental designs in many disciplines (Good, 2005). However, a general overview of the theoretical development and applications of PTs in the educational and behavioral sciences is still lacking, mainly because articles on PTs are spread out over the literature. Accordingly, a systematic review is called for.

The specific objectives of this article are:

1. to provide an overview of the theoretical development of PTs and summarize several key areas of theoretical research;

2. to summarize active areas of educational and behavioral research applying PTs;

3. to identify relevant types of analysis under which PTs have been applied; and

4. to offer articles included in this systematic review as exemplars for educational and behavioral researchers to carry out PTs in their own research.

First, we present the methods of our systematic literature search. Next, we summarize the results of the search for research papers in the educational and behavioral journals for the period 1989-2011[1]. In order to trace the theoretical development of PTs as well as to summarize their applications in the educational and behavioral research, journal articles

[1] This time frame is arbitrary. We started this review in 2009 and wanted to give a 20-year overview. While we were working on the review, additional interesting publications appeared and we added them in the review until 2011.

identified in the search are divided into two categories: theoretical articles and application articles. For the theoretical articles, we identified several key research areas. For the application papers, we identified several active application areas in the educational and behavioral sciences as well as several relevant types of analysis under which PTs are applied. Finally, we discuss the implications of our results and give several recommendations for future research.

Methodology

Systematic search process

We searched ERIC and Web of Science to identify the relevant literature. More specifically, we aimed to include in our study all articles on permutation or randomization tests, or reporting on studies using these tests, that are published between 1989 and 2011 in English-language journals in the field of educational and behavioral sciences[2]. The Appendix presents the details of our search strategy.

Study selection

The selection, based on title and abstract, was done by two review-authors. Articles that do not focus on PTs, are not published in journals, or use PTs in other disciplines were identified and removed from the search results. The two review-authors independently assessed those articles for inclusion. Inter rater agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements. The inter rater reliability was 97.82%. Disagreements were resolved by discussion.

Afterwards, all selected articles were classified into two groups: a theoretical group and an application group. In this way, we intend to summarize the theoretical evolution of

[2] The identification of whether a journal is an educational or behavioral journal is achieved by looking at the journals in the ISI categories related to educational and behavioral sciences or to look at the aims and scope of the journals.

PTs, and the active areas and relevant types of analysis under which these methods have been applied.

In order to make the correct classification, all selected articles were read completely with the following questions in mind:

- Is the article focusing on the theoretical development of PTs? If so, which aspects of PTs are discussed?
- Is the article an application paper? If so, what is its research area in the educational or behavioral research field? In addition, the “data analysis” sections were scrutinized in order to determine the types of analysis under which PTs were carried out.

Data extraction and management

Data were extracted on:

- Author(s), year of publication, and journal
- Classification of articles (theoretical or applied)
- Categories of studies (for theoretical articles)
- Applied areas (for applied articles)
- Type of analysis (for applied articles)

Results

General description of search results

Our search strategy yielded 91 published articles from ERIC and 460 published articles from Web of Science. After removing the duplicate articles and excluding the articles that do not correspond to our inclusion criteria, we obtained 224 published articles (see Figure 1): 141 theoretical and 83 application articles.

Insert Figure 1 about here

The concrete results of the systematic review are illustrated for the “theoretical” and “application” papers separately.

Classification of theoretical papers

The 141 theoretical papers were grouped into six categories. The author(s) and publication year of the retrieved articles for each category are presented in Table 1. The number and percentage of articles for each category are presented in Table 2.

Insert Table 1 about here

Insert Table 2 about here

In the following, we summarize each category in more detail.

Articles introducing PTs. Papers belonging to this category serve pedagogical and introductory purposes. Although PTs are only discussed at an elementary level in these articles (except for Hunter & May, 2003), they play an important role in propagating PTs and their future applications in different disciplines. In many research articles, authors state that classical statistical methods such as two sample *t*-tests are often routinely applied, although in some situations they are problematic, for instance when parametric assumptions (e.g., normally distributed population) are violated. PTs are presented as good alternatives to parametric statistics. Papers in this category can acquaint the educational and behavioral researchers with some basic knowledge of PTs through concrete examples. At the same time, these didactical papers stress the importance of PTs and advocate integrating PTs into the traditional statistical curriculum.

Articles on algorithms, programs and software packages for PTs. The articles in this category show the methodological improvement of PTs in the educational and behavioral sciences. For example, Bakeman et al. (1996) introduced an algorithm about how to use Monte Carlo PTs for the analysis of behavior sequences. Hemelrijk (1990) proposed an algorithm on the use of PTs for the analysis of social interaction matrices, which can be applied to animal behavior analysis. Linting et al. (2011) introduced a technique on how to use PTs for components selection in principle component analysis (PCA). Long et al. (2010) proposed a PT to obtain probability values for chi-squared and likelihood-ratio test statistics for multiway contingency tables. Rounds et al. (1992) introduced PTs for evaluating vocational interest structural hypotheses. Besides new algorithms, articles in this section offer several ways of carrying out PTs through computers. These includes: (1) free available FORTRAN and R programs for specific applications (e.g., Berry & Mielke, 1996; Bulté & Onghena, 2008, 2009; Edgington & Khuller, 1992; Johnston et al., 2008), (2) stand-alone software packages for PTs (e.g., May et al., 1989), (3) and programs through general statistical packages, like SAS and SPSS (e.g., Cai, 2006; Chen & Dunlap, 1993; Hayes, 1998; O'Connor, 1999, 2006).

Articles on the performance of PTs. It has been argued that PTs have good operation characteristics as compared to parametric tests and standard nonparametric tests. Over the last two decades, due to the increment of computer processing speed, the amount of studies focusing on the performance of PTs has substantially increased. Our review includes 27 studies examining the performance of PTs. Researchers have compared the performance of PTs with both parametric and nonparametric rank test counterparts through a range of different scenarios. For instance, van den Brink and van den Brink (1989) compared the power of the *t* test, Wilcoxon's test and approximate PTs for the two-sample location problem. Peres-Neto and Olden (2001) evaluated the performance of permuted ANOVA, parametric ANOVA, and Kruskal-Wallis test for the data from behavioral studies. Mewhort (2005) compared the performance of PTs and *F* tests when the normality of error has been violated. Those studies can help researchers to choose the most appropriate tests for their own experimental settings.

Articles on multivariate PTs. All the above mentioned papers discuss univariate PTs and their performance. There is also a need for distribution-free alternatives for parametric multivariate tests (e.g., MANOVAs). Some researchers (e.g., Blair et al., 1994) suggest the use of multivariate PTs (MPTs) as possible alternatives to statistical inference for multivariate designs when the parametric assumptions, such as the sample populations are multivariate normal, cannot be met. We found 11 articles discussing MPTs issues. Most articles present the principles of MPTs (e.g., Blair & Karniski, 1993; Maris, 2004; Mielke & Berry, 1999) and their performance (e.g., Blair et al., 1994; Finch & Davenport, 2009; Yoder et al., 2004). Besides, Pesarin (1990) proposed a nonparametric combination of marginal dependent PTs. Abbate et al. (2001) extended this combination method to multidimensional situations.

Articles on PTs for single-case designs. When researchers design an experiment, a group design is often the first option. Although very useful, group designs have their limitations, for example in areas that involve very specific and rare types of participants (Edgington & Onghena, 2007). In these areas, single-case experimental designs can be a good alternative to group designs. When conducting statistical analyses for single-case designs, parametric statistical tests, such as *t* tests and ANOVAs, are often inappropriate. PTs are a preferred option compared to these parametric statistics. Articles in this category discuss PTs for single-case designs from several aspects. Some articles discuss how PTs can be applied to different single-case experiments, such as ABAB designs (e.g., Manolov & Solanas, 2008; Onghena, 1992), alternating treatments designs (e.g., Onghena & Edgington, 1994) and multiple-baseline designs (e.g., Ferron & Jones, 2006). Meanwhile, some articles focus on the performance (e.g., in terms of statistical power) of PTs for single-case experiments (e.g., Lall & Levin, 2004; Sierra et al., 2005).

Articles on advanced topics. Some other papers discuss theoretical problems of PTs at an advanced level. For instance, Mewhort et al. (2010) introduced a technique on how to apply PTs to factorial designs. Johnston et al. (2007b) discussed the issue on the number of permutations that is necessary to accurately estimate the *p* value. Long et al. (2007) proposed to use PTs for the estimation of the standard error instead of traditional asymptotic estimates. Mielke et al. (2004) examined a method to combine probability values from independent PTs. Takane and Hwang (2002) and Takane et al. (2008) examined PTs in canonical correlation analysis for testing the statistical significance of canonical correlations. Takane and Hwang (2005) discussed how to use PTs for dimensionality selection.

Classification of application papers

Our review includes 83 articles applying PTs in the educational and behavioral research. Table 3 shows some specific features of each application paper with respect to application areas and types of analysis.

Insert Table 3 about here

Most application articles have employed PTs for data analysis in four active areas. In the following, we summarize each area in more detail.

Vocational interests. Vocational interests, as expressions of personality, are a hot topic in psychological research. Holland's (1985) hexagon model is a popular method on investigating the career interests among individuals. In order to examine the fit of the Holland's circular order relations to any similarity or dissimilarity matrix (e.g., a correlation matrix), a PT for hypothesized order relationships (Hubert & Arabie, 1987) has often been applied. For this PT, the randomization distribution is generated through rearranging the rows and columns of the correlation matrix. For each rearrangement, the number of confirmed predictions of hypothesized order relations for this rearranged correlation matrix is calculated. The *p* value is the proportion of arrangements that can yield the same number of hypothesized predictions or more than that from the observed correlation matrix. This PT can be realized on the computer via a FORTRAN program, RANDALL (Tracey, 1997). We retrieved 19 applications of PTs in the evaluation of the circular order relations in vocational interests (e.g., Anderson et al., 1997; Darcy & Tracey, 2007; Hedrih, 2008; Nagy et al., 2010; Šverko, 2008).

Animal behaviors. The study of animal behaviors is another area in which PTs have been widely used in recent years. Among all retrieved articles in this category, a majority of them concentrated on the discussion of association among animals in a social structure. In those association studies, researchers try to test whether a certain preferred and/or avoided companionship can be explained as a random fashion (Bejder et al., 1998; Whitehead, 1999). A test of random association can be performed through the idea of permutation proposed by Manly (1997). Several articles from this review have taken this approach (e.g., Fujita & Higuchi, 2007; Lemons & Sedinger, 2011; Richardson et al., 2010; Sendova-Franks & Franks, 1995a, 1995b). Later, Bejder et al. (1998) described an updated permutation procedure, with a further modification by Whitehead (1999) for testing the randomness of association. Meanwhile, this algorithm can be realized through a computer program, SOCPROG (Whitehead, 2009). This method has been applied in three articles (Durrell et al., 2004; Ramos-Fernandez et al., 2009; Ramp et al., 2010).

Sometimes the examination of animal association involves multivariate data from groups or individuals. For instance, researchers may intend to assess whether or not birds close to one another have songs that are more similar than those pairs that are located further apart geographically (Schnell, Watt, & Douglas, 1985). Under this kind of circumstances, matrix permutation procedures (Hemelrijk, 1990; Mantel, 1967;) were often applied. With this procedure, two square distance or similarity matrices, each representing different characteristics for the same group or individuals, are compared to determine whether there is a statistical correlation between the corresponding elements. The test statistic is the sum of the products of corresponding matrix elements. After randomly reassigning rows and columns of one matrix, the test statistic is re-calculated. This process is repeated many times, and the null distribution can be generated. We found three articles using Mantel's test to test the association among animals (Mitani & Brandt, 1994; Mitani et al., 2000; Payne et al., 2000).

Brain and behavior. The relationship between brain and behavior is a hot study area in the contemporary psychological and behavioral research. The appearance of model imaging equipment such as MRT (magnetic resonance imaging) makes it possible to position certain cognitive activities associated with specific brain regions. As the statistical parametric mapping approach is not free from parametric assumptions, PTs are strongly recommended because they can be free from the distributional assumptions and deal with the multiple comparisons issue. For instance, PTs have been used to explore the early impact of antipsychotics on brain morphology (Chua et al., 2009) and to examine the relationships between regional activation and behavioral performance measures (Ng et al., 2011).

ERP (event-related brain potential) and EEG (electroencephalogram), as alternative tools on brain research, can provide direct, real-time measures of brain activity associated with successive stages of information processing (Kayser et al., 2007). PTs have proven useful for the comparison of topographies because they are free from distributional assumptions and are still valid even if the number of channels is larger than the sample size (Maris, 2004). For instance, PTs have been used to evaluate topographic old/new differences for paired samples (Kayser et al., 2007).

Gene and behavior. The relationship between gene and behavior is another hot study area in psychological and behavioral research. One of the topics in the study of the effect of genes is whether certain genetic markers are associated with relevant behaviors or diseases. The statistical testing of a marker-disease association analysis usually requires the simultaneous testing of multiple hypotheses, requiring an adjustment on the significance values in order to decrease the Type I error rate. PTs can be used to correct for multiple testing. The PT is performed by randomly assigning the genotypes across subjects with the phenotype variables unchanged so that any relationship between genotype and phenotype can be broken up. With each permutation, a chosen test statistic is re-calculated for the phenotypes. The empirical p value can be achieved by comparing the obtained test statistic to all the other test statistics across all data divisions forming the null distribution. Several articles in this review adopted PTs for multiple comparison on gene and behavior research (e.g., Arias-Vasquez et al., 2011; Lerer et al., 2010; Liao et al., 2009).

Types of analysis

In this section, we summarize the major types of analysis under which PTs have been applied in the educational and behavioral research. The number and percentage of each type of analysis is presented in Table 4.

Insert Table 4 about here

The studies on vocational interests often involve evaluating the fit of hypothesized order relations. By applying PTs of hypothesized order relatons, confirmatory examinations on order relations can be realized. In animal behavior research, PTs have been used to test the existence of social structures. With multivariate data for testing association of social structure, Mantel's randomization tests were often used to examine the correlation between two (or more) square distance matrices.

PTs have also been applied in multivariate designs. Relying on fewer assumptions, MPTs can be used in settings where the parametric MANOVA technique is not feasible (e.g., Kayser et al., 2007; Knyazeva et al., 2010). Moreover, providing more power than the classical Bonferroni method, multiple comparisons corrected by PTs have been widely employed in many studies on the relationship between brain and behavior as well as gene and behavior.

As mentioned above, the group approach of experimental designs is not always possible and single-case experimental designs might be good alternatives. PTs are particularly useful for testing the statistical significance in those single-case studies (e.g., Bardon et al., 2008; Dennin & Ellis, 2003; Huang & Chao, 1999; Koene, 1996; Regan et al., 2005; Ross & Begeny, 2011).

In addition, PTs can be employed in independent two-group designs (e.g., Greeff & Conradie, 1998; Longenecker et al., 2010; Woodruff et al., 1997). Besides, PTs have also been applied under paired two-group designs (e.g., Boltri et al., 2003; Snyder et al., 2010). In a paired two-group design, subjects are tested before and after a treatment. Each measurement of the pretest is linked to one measurement of the posttest. This approach is valuable, particularly when independent two-group experiments are not feasible.

Discussion

Parametric statistical methods, such as standard *t* or *F* tests, are frequently used for hypothesis testing in educational and behavioral research. However, these tests require that the observations are randomly drawn from a normally distributed population, that the population variances are equal, and that the observations are independent of each other. An advantage of PTs is that they do not make any assumptions regarding the probability distribution underlying the data at hand. Additionally, PTs are very easy to apply and versatile, so that researchers can develop a PT for their own particular design (Edgington & Onghena, 2007).

Compared with classical parametric statistical methods, PTs are not frequently used by researchers. One of the reasons for this is probably that most statistical textbooks do not cover this topic. To further advance the development and applications of PTs, an introduction of PTs' rationale and advocating instruction of PTs in basic statistical training are desired. This review reveals that Lock and Lock (2008), Rossman (2008), and Tintle et al. (2011) have made initial efforts to introduce PTs in statistics education. Rodger's (1999) taxonomy gave a good introduction on how to teach sampling. However, more efforts needs to be done to make PTs a standard part like parametric statistical test in statistical curriculum.

It is fortunate to see from this review that the computing burden of PTs is reduced, not only by technological developments, but also by recent efforts of educational and behavioral researchers. For example, Mewhort et al. (2010) introduced a PT for a 2 by 2 factorial design, which decreases the computational load to a large extent. It is noted that the method of Mewhort et al. (2010) is based on Gill's (2007) algorithm. This algorithm can greatly reduce the computing load by using a Fourier expansion to count cases that are as extreme as, or more extreme than, the data observed. Besides new algorithms, educational and behavioral researchers have made their effort to popularize PTs by developing computer software and programs. These includes: (a) basic programming language such as FORTRAN (e.g., Berry & Mielke, 1996, 1998; Edgington & Khuller, 1992); (2) statistical programming language such as SAS (e.g., Chen & Dunlap, 1993), SPSS (e.g., Cai, 2006; Hayes, 1998), and R (Bulté & Onghena, 2008, 2009; Eudey et al., 2010); and (3) stand-alone software (e.g., Dixon et al., 1998; May et al., 1989; McKenzie et al., 1997; Onghena & May, 1995). However, it is unfortunate to see that most of the software and programs perform PTs under univariate designs and the facility to deal with multivariate PTs is still lacking. It is also noted that we are happy to see that parallel computing technique combined with high performance computing has begun to be employed in genetic research (Petrou et al., 2011). Although it only allows for the applications of PTs for multiple comparisons, we regard it as a revolution for PTs.

Meanwhile, methodological possibilities of PTs have been extended over the last 20 years. Research foci on PTs have shifted from basic rationale introduction to more advanced theories, from univariate to multivariate settings, and from simple applications to the possibilities of more complicated applications. Among those articles, a majority of them focused on algorithms and programs of PTs in different contexts, PTs for single-case designs, and performance of PTs for group designs, whereas multivariate PTs were less discussed. Given the theoretical development, PTs have been employed in several active areas of educational and behavioral research and under certain types of experimental designs. It is exciting to see that PTs have begun to be applied in brain and behavior as well as gene and behavior research. In these two research fields, PTs were used to yield the corrected *p* values for multiple comparisons (e.g., Lerer et al., 2010; Liao et al., 2009; Savitz et al., 2008).

However, we also found that PTs were used under simple contexts (e.g., order relations; randomness analysis) in nearly half of the application articles (41 articles in total). To date, the applications of PTs in more complex contexts (e.g., multivariate designs) are still quite limited (3 articles in total).

As to the association between the theoretical and application articles, some applications of PTs can be attributed to the theoretical developments published in the educational and behavioral journals. In the area of vocational interests, PTs of hypothesized order relations have been applied to evaluate the fit of circular order models. The rationale and procedure of this PT were proposed by Hubert and Arabie (1987). This paper, although not included in our review due to its early publishing date, was published in a psychological journal. In the area of brain and behavior, MPTs have been used to make comparisons among topographies. The methods used in the articles were originated from the work by Pesarin (1990), Blair et al. (1994), and Blair and Karniski (1993), which were all covered in the theoretical part. Moreover, the applications of PTs for single-case studies in this review can also be attributed to the theoretical development of PTs discussed in the theoretical section.

Recommendations for future research

Based on the systematic review, we conclude that PTs have received great attention and have been applied in many educational and behavioral studies in the past 20 years but still need to be improved. In the following, we propose several topics that may further promote the development and applications of PTs:

1) More statistical curricula that integrate PTs as part of basic statistical training in the educational and behavioral sciences;
2) Reviews of PTs in other disciplines, so that an overall framework of PTs can be achieved;

3) New algorithms and programs of PTs for two-group, multiple-group, factorial, and multivariate designs based on parallel computing theory;
4) Multivariate PTs software and simulation studies on the performance of multivariate PTs;
5) An R package for PTs (univariate and multivariate PTs);
6) Internet and cloud computing based platform for PTs.

Conclusion

In this review, we discussed the theoretical developments and applications of PTs in educational and behavioral research. On the one hand, we categorized theoretical papers into six categories so that active theoretical developments can be clarified, which indicated the extension of the methodological possibilities of PTs over the past 20 years. On the other hand, main active application areas were summarized. It is good to see that (1) some researchers began to advocate introducing PTs into basic statistics training; (2) computing load for PTs may be reduced dramatically by some intelligent algorithms; (3) PTs began to be applied in new areas such as brain and behavior research and gene and behavior research; (4) besides simple types of analysis such as independent two-group comparison, PTs can also be carried out under more complex situations such as multivariate analysis. All the above information offers one important message: for many (even complex) parametric techniques based on distributional assumptions, alternative PTs can be used. Finally, we offered several recommendations for future research.

References

Abbate, C., Giorgianni, C., Munao, F., Pesarin, F., & Salmaso, L. (2001). Neurobehavioral evaluation in humans exposed to hydrocarbons: A new statistical approach. *Psychotherapy and Psychosomatics, 70*(1), 44–49. doi:10.1159/000056224

Adams, D. C., & Anthony, C. D. (1996). Using randomization techniques to analyse behavioural data. *Animal Behaviour*, *51*, 733–738. doi:10.1006/anbe.1996.0077

Anderson, M. Z., Tracey, T. J. G., & Rounds, J. (1997). Examining the invariance of Holland's vocational interest model across gender. *Journal of Vocational Behavior*, *50*(3), 349–364. doi:10.1006/jvbe.1996.1550

Apostolova, L. G., Lu, P., Rogers, S., Dutton, R. A., Hayashi, K. M., Toga, A. W., Cummings, J. L., et al. (2008). 3D mapping of language networks in clinical and pre-clinical Alzheimer's disease. *Brain and Language*, *104*(1), 33–41. doi:10.1016/j.bandl.2007.03.008

Arias-Vasquez, A., Altink, M. E., Rommelse, N. N. J., Slaats-Willemse, D. I. E., Buschgens, C. J. M., Fliers, E. A., Faraone, S. V., et al. (2011). CDH13 is associated with working memory performance in attention deficit/hyperactivity disorder. *Genes, Brain and Behavior*, *10*(8), 844–851. doi:10.1111/j.1601-183X.2011.00724.x

Avilés, L., Varas, C., & Dyreson, E. (1999). Does the African social spider *Stegodyphus dumicola* control the sex of individual offspring? *Behavioral Ecology and Sociobiology, 46*(4), 237–243. doi:10.1007/s002650050615

Bai, H., & Pan, W. (2008). Resampling methods revisited: Advancing the understanding and applications in educational research. *Journal of Research and Method in Education, 31*(1), 45–62. doi:10.1080/17437270801919909

Bakeman, R., McArthur, D., & Quera, V. (1996). Detecting group differences in sequential association using sampled permutations: Log odds, kappa, and phi compared. *Behavior Research Methods, Instruments, & Computers*, *28*(3), 446–457. doi:10.3758/BF03200524

Bakeman, R., Robinson, B. F., & Quera, V. (1996). Testing sequential association: Estimating exact *p* values using sampled permutations. *Psychological Methods*, *1*(1), 4–15. doi:10.1037//1082-989X.1.1.4

Barbella, P., Denby, L., & Landwehr, J. M. (1990). Beyond eploratory data analysis: The randomization test. *Mathematics Teacher, 83*(2), 144–149.

Bardon, L. A., Dona, D. P., & Symons, F. J. (2008). Extending classwide social skills interventions to at-risk minority students: A preliminary application of randomization tests combined with single-subject design methodology. *Behavioral Disorders*, *33*(3), 141–152.

Bear, G. (1995). Computationally intensive methods warrant reconsideration of pedagogy in statistics. *Behavior Research Methods, Instruments, & Computers*, *27*(2), 144–147. doi:10.3758/BF03204719

Beasley, T. M., Erickson, S., & Allison, D. B. (2009). Rank-based inverse normal transformations are increasingly used, but are they merited? *Behavior Genetics*, *39*(5), 580–595. doi:10.1007/s10519-009-9281-0

Bejder, L., Fletcher, D., & Bräger, S. (1998). A method for testing association patterns of social animals. *Animal Behaviour, 56*, 719–725. doi:10.1006/anbe.1998.0802

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2003). Permutation methods for the analysis of matched-pairs experimental designs. *Psychological Reports*, *92*(3), 1141–1150. doi:10.2466/PR0.92.4.1141-1150

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2005). Resampling probability values for measures of ordinal variation and consensus. *Psychological Reports*, *96*(3), 547–552. doi:10.2466/PR0.96.3.547-552

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2008a). Exact permutation probability values for weighted kappa. *Psychological Reports*, *102*(1), 53–57. doi:10.2466/PRO.102.1.53-57

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2008b). Weighted Kappa for multiple raters. *Perceptual and Motor Skills*, *107*(3), 837–848. doi:10.2466/PMS.107.3.837-848

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2011). Analysis of trend: A permutation alternative to the *F* test. *Perceptual and Motor Skills*, *112*(1), 247–257. doi:10.2466/03.PMS.112.1.247-257

Berry, K. J., Johnston, J. E., Zahran, S., & Mielke, P. W. (2009). Stuart's tau measure of effect size for ordinal variables: Some methodological considerations. *Behavior Research Methods*, *41*(4), 1144–1148. doi:10.3758/BRM.41.4.1144

Berry, K. J., & Mielke, P. W. (1992a). A family of multivariate measures of association for nominal independent variables. *Educational and Psychological Measurement*, *52*(1), 41–55. doi:10.1177/0013164492052001004

Berry, K. J., & Mielke, P. W. (1992b). A measure of association for nominal independent variables. *Educational and Psychological Measurement*, *52*(4), 895–898. doi:10.1177/0013164492052004009

Berry, K. J., & Mielke, P. W. (1994). A test of significance for the index of ordinal variation. *Perceptual and Motor Skills*, *79*(3), 1291–1295.

Berry, K. J., & Mielke, P. W. (1996). Analysis of multivariate matched-paired data: A FORTRAN 77 program. *Perceptual and Motor Skills*, *83*(3), 788–790. doi:10.2466/pms.1996.83.3.788

Berry, K. J., & Mielke, P. W. (1997). Exact and approximate probability values for the Terpstra-Jonckheere test against ordered alternatives. *Perceptual and Motor Skills*, *85*(1), 107–111. doi:10.2466/pms.1997.85.1.107

Berry, K. J., & Mielke, P. W. (1998). A FORTRAN program for permutation covariate analyses of residuals based on Euclidean distance. *Psychological Reports*, *82*(2), 371–375. doi:10.2466/PR0.82.2.371-375

Berry, K. J., Mielke, P. W., & Mielke, H. W. (2002). The Fisher-Pitman permutation test: An attractive alternative to the *F* test. *Psychological Reports*, *90*(2), 495–502. doi:10.2466/PR0.90.2.495-502

Bilker, W. B., Brensinger, C., & Gur, R. C. (2004). A two factor ANOVA-like test for correlated correlations: CORANOVA. *Multivariate Behavioral Research*, *39*(4), 565–594. doi:10.1207/s15327906mbr3904_1

Blair, R. C., Higgins, J. J., Karniski, W., & Kromrey, J. D. (1994). A study of multivariate permutation tests which may replace Hotelling's T^2 test in prescribed circumstances. *Multivariate Behavioral Research*, *29*(2), 141–163. doi:10.1207/s15327906mbr2902_2

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*(5), 518–524. doi:10.1111/j.1469-8986.1993.tb02075.x

Boltri, J. M., Hash, R. B., & Vogel, R. L. (2003). Are family practice residents able to interpret electrocardiograms? *Advances in Health Sciences Education*, *8*(2), 149–153. doi:10.1023/A:1024943613613

Brusco, M. J. (2004). On the concordance among empirical confusion matrices for visual and tactual letter recognition. *Perception & Psychophysics*, *66*(3), 392–397. doi:10.3758/BF03194887

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*(4), 509–540. doi:10.1207/s15327906mbr2704_2

Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. *Behavior Research Methods*, *40*(2), 467–478. doi:10.3758/BRM.40.2.467

Bulté, I., & Onghena, P. (2009). Randomization tests for multiple-baseline designs: An extension of the SCRT-R package. *Behavior Research Methods*, *41*(2), 477–485. doi:10.3758/BRM.41.2.477

Byrne, E., Stillitano, M., Williams, C. J., & Christian, J. C. (2007). *Behavior Genectics, 37*, 617–620. doi:10.1007/s10519-007-9154-3

Cai, L. (2006). Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. *Behavior Research Methods*, *38*(1), 51–59. doi:10.3758/BF03192749

Camilli, G., & Smith, J. (1990). Comparison of the Mantel-Haenszel test with a randomized and a jackknife test for detecting biased items. *Journal of Educational Statistics, 15(1),* 53–67. doi:10.3102/10769986015001053

Campbell, V., Legendre, P., & Lapointe, F.-J. (2009). Assessing congruence among ultrametric distance matrices. *Journal of Classification*, *26*(1), 103–117. doi:10.1007/s00357-009-9028-x

Carlier, M., Beau, J., Marchaland, C., & Michel, F. (1994). Sibling resemblance in two manual laterality tasks. *Neuropsychologia*, *32*(6), 741–746. doi:10.1016/0028-3932(94)90033-7

Chakrabarti, B., Bullmore, E., & Baron-Cohen, S. (2006). Empathizing with basic emotions: Common and discrete neural substrates. *Social Neuroscience*, *1*(3–4), 364–384. doi:10.1080/17470910601041317

Chakrabarti, B., Dudbridge, F., Kent, L., Wheelwright, S., Hill-Cawthorne, G., Allison, C., et al. (2009). Genes related to sex steroids, neural growth, and social-emotional behavior are associated with austic traits, empathy, and Asperger syndrome. *Autism Research*, *2*, 157–177. doi:10.1002/aur.80

Chen, R. S., & Dunlap, W. P. (1993). SAS procedures for approximate randomization tests. *Behavior Research Methods, Instruments, & Computers*, *25*(3), 406–409. doi:10.3758/BF03204532

Chin, L., Haughton, D., & Aczel, A. (1996). Analysis of student evaluation of teaching scores using bootstrap and permutation methods. *Journal of Computing in Higher Education*, *8*(1), 69–84. doi:10.1007/BF02942396

Chua, S. E., Deng, Y., Chen, E. Y. H., Law, C. W., Chiu, C. P. Y., Cheung, C., Wong, J. C. H., et al. (2009). Early striatal hypertrophy in first-episode psychosis within 3 weeks of initiating antipsychotic drug treatment. *Psychological Medicine*, *39*(5), 793–800. doi:10.1017/S0033291708004212

Darcy, M. U. A., & Tracey, T. J. G. (2007). Circumplex structure of Holland’s RIASEC interests across gender and time. *Journal of Counseling Psychology*, *54*(1), 17–31. doi:10.1037/022-0167.54.1.17

Day, S. X., Rounds, J., & Swaney, K. (1998). The structure of vocational interests for diverse racial-ethnic groups. *Psychological Science*, *9*(1), 40–44. doi:10.1111/1467-9280.00007

de Vries, H. (1993). The rowwise correlation between two proximity matrices and the partial rowwise correlation. *Psychometrika*, *58*(1), 53–69. doi:10.1007/BF02294470

de Vries, H. (1995). An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behaviour*, *50*(5), 1375–1389. doi:10.1016/0003-3472(95)80053-0

de Vries, H., Stevens, J. M. G., & Vervaecke, H. (2006). Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, *71*, 585–592. doi:10.1016/j.anbehav.2005.05.015

Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, *72*(4), 563–581. doi:10.1007/s11336-007-9016-1

Dekle, D. J., Leung, D. H. Y., & Zhu, M. (2008). Testing intergroup concordance in ranking experiments with two groups of judges. *Psychological Methods*, *13*(1), 58–71. doi:10.1037/1082-989X.13.1.58

Dennin, M. K., & Ellis, M. V. (2003). Effects of a method of self-supervision for counselor trainees. *Journal of Counseling Psychology*, *50*(1), 69–83. doi:10.1037//0022-0167.50.1.69

Dineen, R. A., Vilisaar, J., Hlinka, J., Bradshaw, C. M., Morgan, P. S., Constantinescu, C. S. et al. (2009). Disconnection as a mechanism for cognitive dysfunction in multiple sclerosis. *Brain*, *132*, 239–249. doi:10.1093/brain/awn275

Dixon, W. E., Woodard, T., & Merry, M. S. (1998). TouchStat: A Monte Carlo program for calculating sequential touching probabilities. *Behavior Research Methods, Instruments, & Computers*, *30*(4), 592–604. doi:10.3758/BF03209476

du Toit, R., & de Bruin, G. P. (2002). The structural validity of Holland's RIASEC model of vocational personality types for young black South African men and women. *Journal of Career Assessment*, *10*(1), 62–77. doi:10.1177/1069072702010001004

Dunlap, W. P. (2003). Some thoughts regarding the sensitivity of *t* to error in estimated within-group variance. *Journal of General Psychology*, *130*(1), 59–65. doi:10.1080/00221300309601274

Dunlap, W. P., Burke, M. J., & Smith-Crowe, K. (2003). Accurate tests of statistical significance for r_{WG} and average deviation interrater agreement indexes. *Journal of Applied Psychology*, *88*(2), 356–362. doi:10.1037/0021-9010.88.2.356

Durrell, J. L., Sneddon, I. A., O'Connell, N. E., & Whitehead, H. (2004). Do pigs form preferential associations? *Applied Animal Behaviour Science*, *89*(1-2), 41–52. doi:10.1016/j.applanim.2004.05.003

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics*, *28*, 181–187.

Dzhafarov, E. N., Ünlü, A., Trendtel, M., & Colonius, H. (2011). Matrices with a given number of violations of regular minimality. *Journal of Mathematical Psychology*, *55*(3), 240–250. doi:10.1016/j.jmp.2010.12.002

Edgington, E. S. (1996). Randomized single-subject experimental designs. *Behaviour Research and Therapy*, *34*(7), 567–574. doi:10.1016/0005-7967(96)00012-5

Edgington, E. S., & Khuller, P. L. V. (1992). A randomization test computer program for trends in repeated-measures data. *Educational and Psychological Measurement*, *52*(1), 93–95. doi:10.1177/001316449205200110

Edgington, E. S., & Onghena, P. (2007). *Randomization tests* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.

Einarsdóttir, S., Rounds, J., Ægisdóttir, S., & Gerstein, L. (2002). The structure of vocational interests in Iceland: Examining Holland's and Gati's RIASEC models. *European Journal of Psychological Assessment*, *18*(1), 85–95. doi:10.1027//1015-5759.18.1.85

Eudey, T., Kerr, J., & Trumbo, B. (2010). Using R to simulate permutation distributions for some elementary experimental designs. *Journal of Statistics Education*, *18*(1). Retrieved from www.amstat.org/publications/jse/v18n1/eudey.pdf

Fehr, T., Wiedenmann, P., & Herrmann, M. (2007). Differences in ERP topographies during color matching of smoking-related and neutral pictures in smokers and non-smokers. *International Journal of Psychophysiology*, *65*(3), 284–293. doi:10.1016/j.ijpsycho.2007.05.006

Ferraro, T., Smith, G., Ballard, D., Zhao, H., Schwebel, C., Gupta, A., Rappaport, E., et al. (2011). Quantitative trait loci for electrical seizure threshold mapped in C57BLKS/J and C57BL/10SnJ mice. *Genes, Brain and Behavior*, *10*(3), 309–315. doi:10.1111/j.1601-183X.2010.00668.x

Ferron, J., & Foster-Johnson, L. (1998). Analyzing single-case data with visually guided randomization tests. *Behavior Research Methods, Instruments, & Computers*, *30*(4), 698–706. doi:10.3758/BF03209489

Ferron, J., Foster-Johnson, L., & Kromrey, J. D. (2003). The functioning of single-case randomization tests with and without random assignment. *Journal of Experimental Education*, *71*(3), 267–288. doi:10.1080/00220970309602066

Ferron, J., & Jones, P. K. (2006). Tests for the visual analysis of response-guided multiple-baseline data. *Journal of Experimental Education*, *75*(1), 66–81. doi:10.3200/JEXE.75.1.66-81

Ferron, J., & Onghena, P. (1996). The power of randomization tests for single-case phase designs. *Journal of Experimental Education*, *64*(3), 231–239.

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. *Journal of Experimental Education*, *70*(2), 165–178. doi:10.1080/00220970209599504

Ferron, J., & Ware, W. (1994). Using randomization tests with responsive single-case designs. *Behaviour research and therapy*, *32*(7), 787–791. doi:10.1016/0005-7967(94)90037-X

Ferron, J., & Ware, W. (1995). Analyzing single-case data: The power of randomization tests. *Journal of Experimental Education*, *63*(2), 167–178. doi:10.1080/00220973.1995.9943820

Finch, W. H., & Davenport, T. (2009). Performance of Monte Carlo permutation and approximate tests for multivariate means comparisons with small sample sizes when parametric assumptions are violated. *Methodology-European Journal of Research Methods for the Behavioral and Social Sciences*, *5*(2), 60–70. doi:10.1027/1614-2241.5.2.60

Fisch, G. S. (2001). Evaluating data from behavioral analysis: visual inspection or statistical models? *Behavioural Processes*, *54*(1-3), 137–154. doi:10.1016/S0376-6357(01)00155-3

Fischhoff, I. R., Dushoff, J., Sundaresan, S. R., Cordingley, J. E., & Rubenstein, D. I. (2009). Reproductive status influences group size and persistence of bonds in male plains zebra (Equus burchelli). *Behavioral Ecology and Sociobiology*, *63*(7), 1035–1043. doi:10.1007/s00265-009-0723-8

Fouad, N. A., Harmon, L. W., & Borgen, F. H. (1997). Structure of interests in employed male and female members of US racial-ethnic minority and nonminority groups. *Journal of Counseling Psychology*, *44*(4), 339–345. doi:10.1037/0022-0167.44.4.339

Franks, N. R., Sendova-Franks, A. B., & Anderson, C. (2001). Division of labour within teams of New World and Old World army ants. *Animal Behaviour*, *62*(4), 635–642. doi:10.1006/anbe.2001.1794

Fujita, G., & Higuchi, H. (2007). Barn swallows prefer to nest at sites hidden from neighboring nests within a loose colony. *Journal of Ethology*, *25*(2), 117–123. doi:10.1007/s10164-006-0005-0

Gill, P. M. W. (2007). Efficient calculation of *p*-values in linear-statistic permutation significance tests. *Journal of Statistical Computation & Simulation*, *77*(1), 55–61. doi:10.1080/10629360500108053

Glidden-Tracey, C. E., & Parraga, M. I. (1996). Assessing the structure of vocational interests among Bolivian university students. *Journal of Vocational Behavior*, *48*(1), 96–106. doi:10.1006/jvbe.1996.0010

Glidden-Tracey, C., & Greenwood, A. K. (1997). A validation study of the Spanish self-directed search using back-translation procedures. *Journal of Career Assessment*, *5*(1), 105–113. doi:10.1177/106907279700500107

Goldstein, N. E., Arnold, D. H., Rosenberg, J. L., Stowe, R. A., & Ortiz, C. (2001). Contagion of aggression in day care classrooms as a function of peer and teacher responses. *Journal of Educational Psychology*, *93*(4), 708–719. doi:10.1037//0022-0663.93.4.708

Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, *42*(1), 23–28. doi:10.3758/BRM.42.1.23

Good, P. A. (2005). *Permutation tests: A practical guide to resampling methods for testing hypotheses* (3rd ed.). New York: Springer-Verlag.

Greeff, A. P., & Conradie, W. S. (1998). Use of progressive relaxation training for chronic alcoholics with insomnia. *Psychological Reports*, *82*(2), 407–412. doi:10.2466/PR0.82.2.407-412

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *48*(12), 1726–1737. doi:10.1111/j.1469-8986.2011.01272.x

Gutierrez-Garcia, J. M., & Tusell, F. (1997). Suicides and the lunar cycle. *Psychological Reports*, *80*(1), 243–250. doi:10.2466/PR0.80.1.243-250

Haardörfer, R., & Gagné, P. (2010). The use of randomization tests in single-subject research. *Focus on Autism and Other Developmental Disabilities*, *25*(1), 47–54. doi:10.1177/1088357609350745

Hahn, M. K., Blackford, J. U., Haman, K., Mazei-Robison, M., English, B. A., Prasad, H. C., Steele, A., et al. (2008). Multivariate permutation analysis associates multiple polymorphisms with subphenotypes of major depression. *Genes, Brain and Behavior*, *7*(4), 487–495. doi:10.1111/j.1601-183X.2007.00384.x

Hale, J., Reddy, L., Semrud-Clikeman, M., Hain, L., Whitaker, J., Morley, J., Lawrence, K., et al. (2011). Executive impairment determines ADHD medication response: implications for academic achievement. *Journal of Learning Disabilities*, *44*(2), 196–212. doi:10.1177/0022219410391191

Hayes, A. F. (1996). Permutation test is not distribution-free: Testing H_0: ρ=0. *Psychological Methods*, *1*(2), 184–198. doi:10.1037/1082-989X.1.2.184

Hayes, A. F. (1998). SPSS procedures for approximate randomization tests. *Behavior Research Methods, Instruments, & Computers*, *30*(3), 536–543. doi:10.3758/BF03200687

Hayes, A. F., & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. *British Journal of Mathematical and Statistical Psychology*, *60*(2), 217–244. doi:10.1348/000711005X62576

Hedrih, V. (2008). Structure of vocational interests in Serbia: Evaluation of the spherical model. *Journal of Vocational Behavior*, *73*(1), 13–23. doi:10.1016/j.jvb.2007.12.004

Hemelrijk, C. K. (1990). Models of, and tests for, reciprocity, unidirectionality and other social interaction patterns at a group level. *Animal Behaviour*, *39*(6), 1013–1029. doi:10.1016/S0003-3472(05)80775-4

Hofsess, C. D., & Tracey, T. J. G. (2005). The interpersonal circumplex as a model of interpersonal capabilities. *Journal of Personality Assessment*, *84*(2), 137–147. doi:10.1207/s15327752jpa8402_03

Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliffs, NJ: Prentice–Hall.

Huang, J. Y., & Chao, L. (1999). Specific intervention for multiplication skill with learning-disabled students. *Psychological Reports*, *84*(2), 662–662. doi:10.2466/PR0.84.2.662-662

Hubert, L. J., & Arabie, P. (1987). Evaluating order hypotheses within proximity matrices. *Psychological Bulletin, 102*, 172–178. doi:10.1037/0033-2909.102.1.172

Hunter, M., & May, R. (2003). Statistical testing and null distributions: What to do when samples are not random. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale*, *57*(3), 176–188. doi:10.1037/h0087424

Johnston, J. E., Berry, K. J., & Mielke, P. W. (2007a). Resampling permutation probability values for six measures of qualitative variation. *Perceptual and Motor Skills*, *104*(3), 773–776. doi:10.2466/PMS.104.3.773-776

Johnston, J. E., Berry, K. J., & Mielke, P. W. (2007b). Permutation tests: Precision in estimating probability values. *Perceptual and Motor Skills*, *105*(3), 915–920. doi:10.2466/PMS.105.3.915-920

Johnston, J. E., Berry, K. J., & Mielke, P. W. (2008). Resampling permutation probability values or weighted kappa. *Psychological Reports*, *103*(2), 467–475. doi:10.2466/PR0.103.2.467-475

Kayser, J., Tenke, C. E., Gates, N. A., & Bruder, G. E. (2007). Reference-independent ERP old/new effects of auditory and visual word recognition memory: Joint extraction of stimulus-and response-locked neuronal generator patterns. *Psychophysiology*, *44*(6), 949–967. doi:10.1111/j.1469-8986.2007.00562.x

Kempthorne, O., & Doerfler, T. E. (1969) The behavior of some significance tests under experimental randomization. *Biometrika*, *56*, 231–248. doi:10.2307/2334417

Kennedy, M., Spencer, H. G., & Gray, R. D. (1996). Hop, step and gape: Do the social displays of the Pelecaniformes reflect phylogeny? *Animal Behaviour, 51*, 273–291. doi:10.1006/anbe.1996.0028

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, *44*(4), 592–617. doi:10.1006/jmla.2000.2745

Kimberg, D. Y., Coslett, H. B., & Schwartz, M. F. (2007). Power in voxel-based lesion-symptom mapping. *Journal of Cognitive Neuroscience*, *19*(7), 1067–1080. doi:10.1162/jocn.2007.19.7.1067

Kline, T. J. B. (1991). A study of job satisfaction: making a case for using small sample sizes, including a variety of variables and testing for significance with randomized tests. *College and University*, *67*(1), 15–22.

Knyazeva, M. G., Carmeli, C., Fornari, E., Meuli, R., Small, M., Frackowiak, R. S., & Maeder, P. (2010). Binding under conflict conditions: state–space analysis of multivariate EEG synchronization. *Journal of Cognitive Neuroscience*, *23*(9), 2363–2375. doi:10.1162/jocn.2010.21588

Koene, P. (1996). Temporal structure of red jungle fowl crow sequences: single-case analysis. *Behavioural Processes, 38*(2), 193–202. doi:10.1016/S0376-6357(96)00028-9

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychological Methods*, *15*(2), 124–144. doi:10.1037/a0017736

Krause, S., Mattner, L., James, R., Guttridge, T., Corcoran, M. J., Gruber, S. H., & Krause, J. (2009). Social network analysis and valid Markov chain Monte Carlo tests of null models. *Behavioral Ecology and Sociobiology*, *63*(7), 1089–1096. doi:10.1007/s00265-009-0746-1

LaFleur, B. J., & Greevy, R. A. (2009). Introduction to permutation and resampling-based hypothesis tests. *Journal of Clinical Child & Adolescent Psychology*, *38*(2), 286–294. doi:10.1080/15374410902740411

Lall, V. F., & Levin, J. R. (2004). An empirical investigation of the statistical properties of generalized single-case randomization tests. *Journal of School Psychology*, *42*(1), 61–86. doi:10.1016/j.jsp.2003.11.002

Lapointe, F. J., & Garland, T. (2001). A generalized permutation model for the analysis of cross-species data. *Journal of Classification*, *18*(1), 109–127.

Lapointe, F. J., & Legendre, P. (1995). Comparison tests for dendrograms: A comparative evaluation. *Journal of Classification*, *12*(2), 265–282.

Legendre, P., & Lapointe, F. J. (1995). Matching behavioral evolution to brain morphology. *Brain, Behavior and Evolution*, *45*(2), 110–121. doi:10.1159/000113544

Lehmann, E. L., & D'Abrera, H. J. (2006). *Nonparametrics: Statistical methods based on ranks* (2nd ed.). New York, NY: Springer.

Lemons, P. R., & Sedinger, J. S. (2011). Egg size matching by an intraspecific brood parasite. *Behavioral Ecology*, *22*(4), 696–700. doi:10.1093/beheco/arr035

Leong, F. T. L., Austin, J. T., Sekaran, U., & Komarraju, M. (1998). An evaluation of the cross-cultural validity of Holland’s theory: Career choices by workers in India. *Journal of Vocational Behavior*, *52*(3), 441–455. doi:10.1006/jvbe.1997.1637

Lerer, E., Levi, S., Israel, S., Yaari, M., Nemanov, L., Mankuta, D., Yirmiya, N., et al. (2010). Low CD38 expression in Lymphoblastoid cells and Haplotypes are both associated with Autism in a family-based study. *Autism Research*, *3*(6), 293–302. doi:10.1002/aur.156

Levin, J. R., & Wampold, B. E. (1999). Generalized single-case randomization tests: Flexible analyses for a variety of situations. *School Psychology Quarterly*, *14*(1), 59–93. doi:10.1037/h0088998

Levin, J. R., Lall, V. F., & Kratochwill, T. R. (2011). Extensions of a versatile randomization test for assessing single-case intervention effects. *Journal of School Psychology*, *49*(1), 55–79. doi:10.1016/j.jsp.2010.09.002

Li, H. H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika*, *61*(4), 647–677. doi:10.1007/BF02294041

Liao, S.-Y., Lin, S.-H., Liu, C.-M., Hsieh, M. H., Hwang, T. J., Liu, S. K., Guo, S.-C., et al. (2009). Genetic variants in COMT and neurocognitive impairment in families of patients with schizophrenia. *Genes, Brain and Behavior*, *8*(2), 228–237. doi:10.1111/j.1601-183X.2008.00467.x

Linting, M., van Os, B. J., & Meulman, J. J. (2011). Statistical significance of the contribution of variables to the PCA solution: An alternative permutation strategy. *Psychometrika*, *76*(3), 1–21. doi:10.1007/s11336-011-9216-6

Lix, L. M., & Sajobi, T. (2010). Testing multiple outcomes in repeated measures designs. *Psychological Methods*, *15*(3), 268–280. doi:10.1037/a0017737

Lock, R., & Lock, P. (2008). Introducing statistical inference to biology students through bootstrapping and randomization. *PRIMUS, 18*(1), 39–48. doi:10.1080/10511970701745015

Long, L., & Tracey, T. J. G. (2006). Structure of RIASEC scores in China: A structural meta-analysis. *Journal of Vocational Behavior*, *68*(1), 39–51. doi:10.1016/j.jvb.2005.01.002

Long, M. A., Berry, K. J., & Mielke, P. W. (2007). A note on permutation tests of significance for multiple regression coefficients. *Psychological Reports*, *100*(2), 339–345. doi:10.2466/PR0.100.2.339-345

Long, M. A., Berry, K. J., & Mielke, P. W. (2009). Tetrachoric correlation: A permutation alternative. *Educational and Psychological Measurement*, *69*(3), 429–437. doi:10.1177/0013164408324463

Long, M. A., Berry, K. J., & Mielke, P. W. (2010). Multiway contingency tables: Monte Carlo resampling probability values for the chi-squared and likelihood-ratio tests. *Psychological Reports*, *107*(2), 501–510. doi:10.2466/03.PR0.107.5.501-510

Longenecker, J., Kohn, P., Liu, S., Zoltick, B., Weinberger, D. R., & Elvevag, B. (2010). Data-driven methodology illustrating mechanisms underlying word list recall: Applications to clinical research. *Neuropsychology*, *24*(5), 625–636. doi:10.1037/a0019368

Ludbrook, J. (1994). Advantages of permutation (randomization) tests in clinical and experimental pharmacology and physiology. *Clinical & Experimental Pharmacology & Physiology*, *21*, 673–686. doi:10.1111/j.1440-1681.1994.tb02570.x

Lunneborg, C. E. (2001). Random assignment of available cases: Bootstrap standard errors and confidence intervals. *Psychological Methods*, *6*(4), 402–412. doi:10.1037//1082-989X.6.4.402

Maguire, T., & Rogers, W. (1989). Proposed solutions for nonrandomness in educational research. *Canadian Journal of Education/Revue canadienne de l'education, 14*(2), 170–181.

Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.). New York: Chapman & Hall.

Manolov, R., & Solanas, A. (2008). Randomization tests for ABAB designs: Comparing-data-division-specific and common distributions. *Psicothema, 20*(2), 297–303.

Manolov, R., & Solanas, A. (2009). Problems of the randomization test for AB designs. *Psicologica*, *30*(1), 137–154.

Manolov, R., Solanas, A., Bulte, I., & Onghena, P. (2010). Data-division-specific robustness and power of randomization tests for ABAB designs. *Journal of Experimental Education*, *78*(2), 191–214. doi:10.1080/00220970903292827

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, *41*(1), 142–151. doi:10.1111/j.1469-8986.2003.00139.x

May, R. B., & Hunter, M. A. (1993). Some advantages of permutation tests. *Canadian Psychology/Psychologie canadienne, 34*(4), 401–407. doi:10.1037/h0078862

May, R. B., Masson, M. E. J., & Hunter, M. A. (1989). Randomization tests: Viable alternatives to normal curve tests. *Behavior Research Methods, Instruments, & Computers*, *21*(4), 482–483. doi:10.3758/BF03202823

McKenzie, D. P., Mackinnon, A. J., & David, M. C. (1997). KAPCOM: a program for the comparison of kappa coefficients obtained from the same sample of observations. *Perceptual and Motor Skills*, *85*(3), 899–902. doi:10.2466/pms.1997.85.3.899

McKenzie, D. P., Onghena, P., Hogenraad, R., Martindale, C., & MacKinnon, A. J. (1999). Detecting patterns by one-sample runs test: Paradox, explanation, and a new omnibus procedure. *Journal of Experimental Education, 67*(2), 167–179. doi:10.1080/00220979909598351

Medina, J., Kimberg, D. Y., Chatterjee, A., & Coslett, H. B. (2010). Inappropriate usage of the Brunner-Munzel test in recent voxel-based lesion-symptom mapping studies. *Neuropsychologia*, *48*(1), 341–343. doi:10.1016/j.neuropsychologia.2009.09.016

Medland, S. E., Schmitt, J. E., Webb, B. T., Kuo, P.-H., & Neale, M. C. (2009). Efficient calculation of empirical p-values for genome-wide linkage analysis through weighted permutation. *Behavior Genetics*, *39*(1), 91–100. doi:10.1007/s10519-008-9229-9

Mewhort, D. J. K. (2005). A comparison of the randomization test with the *F* test when error is skewed. *Behavior Research Methods*, *37*(3), 426–435. doi:10.3758/BF03192711

Mewhort, D. J. K., Johns, B. T., & Kelly, M. (2010). Applying the permutation test to factorial designs. *Behavior Research Methods*, *42*(2), 366–372. doi:10.3758/BRM.42.2.366

Mewhort, D. J. K., Kelly, M., & Johns, B. T. (2009). Randomization tests and the unequal-N/unequal-variance problem. *Behavior Research Methods*, *41*(3), 664–667. doi:10.3758/BRM.41.3.664

Mielke, P. W., & Berry, K. J. (1994). Permutation tests for common locations among samples with unequal variances. *Journal of Educational and Behavioral Statistics*, *19*(3), 217–236. doi:10.3102/10769986019003217

Mielke, P. W., & Berry, K. J. (1999). Multivariate tests for correlated data in completely randomized designs. *Journal of Educational and Behavioral Statistics*, *24*(2), 109–131. doi:10.2307/1165197

Mielke, P. W., & Berry, K. J. (2000). The Terpstra-Jonckheere test for ordered alternatives: Randomized probability values. *Perceptual and Motor Skills*, *91*(2), 447–450. doi:10.2466/PMS.91.6.447-450

Mielke, P. W., & Berry, K. J. (2002). Data-dependent analyses in psychological research. *Psychological Reports*, *91*(3), 1225–1234. doi:10.2466/PR0.91.8.1225-1234

Mielke, P. W., & Berry, K. J. (2007a). Two-sample multivariate similarity permutation comparison. *Psychological Reports*, *100*(1), 257–262. doi:10.2466/PR0.100.1.257-262

Mielke, P. W., & Berry, K. J. (2007b). *Permutation methods: A distance function approach* (2nd ed.). New York: Spring-Verlag.

Mielke, P. W., Berry, K. J., & Johnston, J. E. (2004). Asymptotic log-linear analysis: Some cautions concerning sparse frequency tables. *Psychological Reports*, *94*(1), 19–32. doi:10.2466/pr0.94.1.19-32

Mielke, P. W., Berry, K. J., & Neidt, C. O. (1996). Permutation test for multivariate matched-pair analyses: Comparisons with Hotelling's multivariate matched-pair T^2 test. *Psychological Reports*, *78*(3), 1003–1008.

Mielke, P. W., Johnston, J. E., & Berry, K. J. (2004). Combining probability values from independent permutation tests: A discrete analog of Fisher's classical method. *Psychological Reports*, *95*(2), 449–458. doi:10.2466/pr0.95.2.449-458

Miller, P., Shapiro, A., Tyack, P., & Solow, A. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, Orcinus orca. *Animal Behaviour*, *67*, 1099–1107. doi:10.1016/j.anbehav.2003.06.017

Mitani, J. C., & Amsler, S. J. (2003). Social and spatial aspects of male subgrouping in a community of wild chimpanzees. *Behaviour*, *140*, 869–884. doi:10.1163/156853903770238355

Mitani, J. C., & Brandt, K. L. (1994). Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, *96*(3), 233–252. doi:10.1111/j.1439-0310.1994.tb01012.x

Mitani, J. C., Merriwether, D. A., & Zhang, C. B. (2000). Male affiliation, cooperation and kinship in wild chimpanzees. *Animal Behaviour*, *59*, 885–893. doi:10.1006/anbe.1999.1389

Mundry, R. (1999). Testing related samples with missing values: A permutation approach. *Animal Behaviour*, *58*(5), 1143–1153. doi:10.1006/anbe.1999.1246

Mundry, R., & Sommer, C. (2007). Discriminant function analysis with nonindependent data: consequences and an alternative. *Animal Behaviour*, *74*, 965–976. doi:10.1016/j.anbehav.2006.12.028

Myors, B. (1998). A Monte Carlo comparison of three tests of the Holland hexagon. *Journal of Vocational Behavior*, *53*(2), 215–226. doi:10.1006/jvbe.1998.1649

Nagy, G., Trautwein, U., & Lüdtke, O. (2010). The structure of vocational interests in Germany: Different methodologies, different conclusions. *Journal of Vocational Behavior*, *76*(2), 153–169. doi:10.1016/j.jvb.2007.07.002

Neuhäuser, M., & Manly, B. F. J. (2004). The Fisher-Pitman permutation test when testing for differences in mean and variance. *Psychological Reports*, *94*(1), 189–194. doi:10.2466/pr0.94.1.189-194

Neuhäuser, M., & Ruxton, G. D. (2009). Distribution-free two-sample comparisons in the case of heterogeneous variances. *Behavioral Ecology and Sociobiology*, *63*(4), 617–623. doi:10.1007/s00265-008-0683-4

Ng, V. W. K., Cooper, A., Williams, S. C. R., Bullmore, E. T., de Zubicaray, G. I., & Suckling, J. (2001). Identifying rate-limiting nodes in large-scale cortical networks for visuospatial processing: An illustration using fMRI. *Journal of Cognitive Neuroscience*, *13*(4), 537–545.

Normando, S., Corain, L., Salvadoretti, M., Meers, L., & Valsecchi, P. (2009). Effects of an enhanced human interaction program on shelter dogs' behaviour analysed using a novel nonparametric test. *Applied Animal Behaviour Science*, *116*(2-4), 211–219. doi:10.1016/j.applanim.2008.10.005

O'Connor, B. P. (1999). Simple and flexible SAS and SPSS programs for analyzing lag-sequential categorical data. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 718–726. doi:10.3758/BF03200753

O'Connor, B. P. (2006). Programs for problems created by continuous variable distributions in moderated multiple regression. *Organizational Research Methods*, *9*(4), 554–567. doi:10.1177/1094428106286984

Oden, N. L., & Sokal, R. R. (1992). An investigation of three-matrix permutation tests. *Journal of Classification*, *9*(2), 275–290. doi:10.1007/BF02621410

Onghena, P. (1992). Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral Assessment*, *14*, 153–153.

Onghena, P., & Edgington, E. S. (1994). Randomization tests for restricted alternating treatments designs. *Behaviour Research and Therapy*, *32*(7), 783–786. doi:10.1016/0005-7967(94)90036-1

Onghena, P., & May, R. B. (1995). Pitfalls in computing and interpreting randomization test *p* values: A commentary on Chen and Dunlap. *Behavior Research Methods, Instruments, & Computers*, *27*(3), 408–411. doi:10.3758/BF03200438

Payne, R. B., Woods, J. L., Siddall, M. E., & Parr, C. S. (2000). Randomization analyses: Mimicry, geographic variation and cultural evolution of song in brood-parasitic straw-tailed whydahs, Vidua fischeri. *Ethology*, *106*(3), 261–282. doi:10.1046/j.1439-0310.2000.00528.x

Perdices, M., & Tate, R. L. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation*, *19*(6), 904–927. doi:10.1080/09602010903040691

Peres-Neto, P. R., & Olden, J. D. (2001). Assessing the robustness of randomization tests: examples from behavioural studies. *Animal Behaviour*, *61*, 79–86. doi:10.1006/anbe.2000.1576

Pesarin, F. (1990). On a nonparametric combination method for dependent permutation tests with applications. *Psychotherapy and Psychosomatics*, *54*(2-3), 172–179. doi:10.1159/000100210

Petrou, S., Sloan, T. M., Mewissen, M., Forster, T., Piotrowski, M., Dobrzelecki, B., Ghazal, P., et al. (2011). Optimization of a parallel permutation testing function for the SPRINT R package. *Concurrency and Computation: Practice and Experience*, *23*(17), 2258–2268. doi:10.1002/cpe.1787

Prelog, A. J., Berry, K. J., & Mielke, P. W. (2009). Resampling permutation probability values for Cronbach's Alpha. *Perceptual and Motor Skills*, *108*(2), 431–438. doi:10.2466/PMS.108.2.431-438

Ramos-Fernandez, G., Boyer, D., Aureli, F., & Vick, L. G. (2009). Association networks in spider monkeys (Ateles geoffroyi). *Behavioral Ecology and Sociobiology*, *63*(7), 999–1013. doi:10.1007/s00265-009-0719-4

Ramp, C., Hagen, W., Palsbøll, P., Bérubé, M., & Sears, R. (2010). Age-related multi-year associations in female humpback whales (Megaptera novaeangliae). *Behavioral Ecology and Sociobiology*, *64*(10), 1563–1576. doi:10.1007/s00265-010-0970-8

Regan, K. S., Mastropieri, M. A., & Scruggs, T. E. (2005). Promoting expressive writing among students with emotional and behavioral disturbance via dialogue journals. *Behavioral Disorders*, *31*(1), 33–50.

Ribi, G., & Porter, A. H. (1995). Mating between two hybridizing species, Viviparus ater and V. contectus (Mollusca: Prosobranchia). *Animal Behaviour*, *49*(5), 1389–1398. doi:10.1006/anbe.1995.0169

Richardson, K., Ewen, J. G., Armstrong, D. P., & Hauber, M. E. (2010). Sex-specific shifts in natal dispersal dynamics in a reintroduced hihi population. *Behaviour*, *147*(12), 1517–1532. doi:10.1163/000579510X531284

Ridler, K., Bullmore, E. T., De Vries, P. J., Suckling, J., Barker, G. J., Meara, S. J. P., Williams, S. C. R., et al. (2001). Widespread anatomical abnormalities of grey and white matter structure in tuberous sclerosis. *Psychological Medicine*, *31*(08), 1437–1446. doi:10.1017/S0033291701004561

Roberts, G., & Evans, P. R. (1993). A method for the detection of non-random associations among flocking birds and its application to sanderlings Calidris alba wintering in N.E. England. *Behavioral Ecology and Sociobiology*, *32*(5), 349–354. doi:10.1007/BF00183791

Rodgers, J. L. (1999). The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivariate Behavioral Research*, *34*(4), 441–456. doi:10.1207/S15327906MBR3404_2

Roe, A. (1956). *The psychology of occupations*. New York: Wiley.

Ross, S. G., & Begeny, J. C. (2011). Improving Latino, English language learners' reading fluency: The effects of small-group and one-on-one intervention. *Psychology in the Schools*, *48*(6), 604–618. doi:10.1002/pits.20575

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7*(2), 5–19.

Rounds, J., & Tracey, T. J. G. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology*, *43*(3), 310–329. doi:10.1037/0022-0167.43.3.310

Rounds, J., Tracey, T. J. G., & Hubert, L. (1992). Methods for evaluating vocational interest structural hypotheses. *Journal of Vocational Behavior*, *40*(2), 239–259. doi:10.1016/0001-8791(92)90073-9

Ryan, J. M., Tracey, T. J. G., & Rounds, J. (1996). Generalizability of Holland's structure of vocational interests across ethnicity, gender, and socioeconomic status. *Journal of Counseling Psychology*, *43*(3), 330–337. doi:10.1037//0022-0167.43.3.330

Samuel, C., Louis-Dreyfus, A., Kaschel, R., Makiela, E., Troubat, M., Anselmi, N., Cannizzo, V., et al. (2000). Rehabilitation of very severe unilateral neglect by visuo-spatio-motor cueing: Two single case studies. *Neuropsychological Rehabilitation*, *10*(4), 385–399. doi:10.1080/096020100411970

Savitz, J., van der Merwe, L., & Ramesar, R. (2008). Personality endophenotypes for bipolar affective disorder: a family-based genetic association analysis. *Genes, Brain and Behavior*, *7*(8), 869–876. doi:10.1111/j.1601-183X.2008.00426.x

Schnell, G. D., Watt, D. J., & Douglas, M. E. (1985). Statistical comparison of proximity matrices: applications in animal behaviour. *Animal Behaviour*, *33*, 239–253. doi:10.1016/S0003-3472(85)80138-X

Sendova-Franks, A. B., & Franks, N. R. (1995a). Spatial relationships within nests of the ant Leptothorax unifasciatus(Latr.) and their implications for the division of labour. *Animal Behaviour*, *50*(1), 121–136. doi:10.1006/anbe.1995.0226

Sendova-Franks, A. B., & Franks, N. R. (1995b). Demonstrating new social interactions in ant colonies through randomization tests: separating seeing from believing. *Animal Behaviour*, *50*(6), 1683–1696. doi:10.1016/0003-3472(95)80021-2

Sierra, V., Solanas, A., & Quera, V. (2005). Randomization tests for systematic single-case designs are not always appropriate. *Journal of Experimental Education*, *73*(2), 140–160. doi:10.3200/JEXE.73.2.140-160

Snyder, F., Flay, B., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K.-K. (2010). Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster-randomized, controlled trial. *Journal of Research on Educational Effectiveness, 3*(1), 26–55. doi:10.1080/19345740903353436

Solanas, A., Sierra, V., Quera, V., & Manolov, R. (2008). Random assignment of intervention in two-phase single-case designs: Data-division-specific distributions. *Psychological Reports*, *103*(2), 499–515. doi:10.2466/PR0.103.2.499-515

Spilková, J., & Hochel, M. (2009). Toward the economy of pedestrian movement in Czech and Slovak shopping malls. *Environment and Behavior*, *41*(3), 443–455. doi:10.1177/0013916508325003

Stahl, S. (2004). Permutation-based methods for examining confusion data in ESP experiments. *Journal of Parapsychology*, *68*(2), 381–403.

Sundaresan, S. R., Fischhoff, I. R., & Dushoff, J. (2009). Avoiding spurious findings of nonrandom social structure in association data. *Animal Behaviour*, *77*(6), 1381–1385. doi:10.1016/j.anbehav.2009.01.021

Sunderland, A., Walker, C. M., & Walker, M. F. (2006). Action errors and dressing disability after stroke: An ecological approach to neuropsychological assessment and intervention. *Neuropsychological Rehabilitation*, *16*(6), 666–683. doi:10.1080/09602010500204385

Šverko, I. (2008). Spherical model of interests in Croatia. *Journal of Vocational Behavior, 72*(1), 14–24. doi:10.1016/j.jvb.2007.10.001

Takane, Y., Hwang, H., & Abdi, H. (2008). Regularized multiple-set canonical correlation analysis. *Psychometrika*, *73*(4), 753–775. doi:10.1007/s11336-008-9065-0

Takane, Y., & Hwang, H. S. (2002). Generalized constrained canonical correlation analysis. *Multivariate Behavioral Research*, *37*(2), 163–195. doi:10.1207/S15327906MBR3702_01

Takane, Y., & Hwang, H. S. (2005). On a test of dimensionality in redundancy analysis. *Psychometrika*, *70*(2), 271–281. doi:10.1007/s11336-003-1089-x

Tan, E. S., Imbos, T., & Does, R. J. M. M. (1994). A distribution-free approach for comparing growth of knowledge. *Journal of Educational Measurement*, *31*(1), 51–65. doi:10.1111/j.1745-3984.1994.tb00434.x

Tintle, N., van der Stoep, J., Holmes, V.-L., Quisenberry, B., & Swanson, T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education, 19*(1). Retrieved from www.amstat.org/publications/jse/v19n1/tintle.pdf

Tracey, T. J. G. (1997). RANDALL: A Microsoft FORTRAN program for a randomization test of hypothesized order relations. *Educational and Psychological Measurement*, *57*(1), 164–168. doi:10.1177/0013164497057001012

Tracey, T. J. G., & Rounds, J. (1994). An examination of the structure of Roe's eight interest fields. *Journal of Vocational Behavior, 44*(3), 279–296. doi:10.1006/jvbe.1994.1019

Tracey, T. J. G., & Ward, C. C. (1998). The structure of children's interests and competence perceptions. *Journal of Counseling Psychology*, *45*(3), 290–303. doi:10.1037//0022-0167.45.3.290

Trainer, J. M., McDonald, D. B., & Learn, W. A. (2002). The development of coordinated singing in cooperatively displaying long-tailed manikins. *Behavioral Ecology, 13*(1), 65–69. doi:10.1093/beheco/13.1.65

Utts, J. (1989). Randomness and randomization tests - A reply to Gilmore. *Journal of Parapsychology*, *53*(4), 345–351.

van den Brink, W. P., & van den Brink, S. G. J. (1989). A comparison of the power of the *t* test, Wilcoxon's test, and the approximate permutation test for the two-sample location problem. *British Journal of Mathematical & Statistical Psychology*, *42*(2), 183–189. doi:10.1111/j.2044-8317.1989.tb00907.x

Vera, J. F., Heiser, W. J., & Murillo, A. (2007). Global optimization in any minkowski metric: A permutation-translation simulated annealing algorithm for multidimensional scaling. *Journal of Classification*, *24*(2), 277–301. doi:10.1007/s00357-007-0020-1

Verboon, P. (1993). Robust nonlinear regression analysis. *British Journal of Mathematical and Statistical Psychology*, *46*(1), 77–94. doi:10.1111/j.2044-8317.1993.tb01003.x

Verdolin, J. L., & Slobodchikoff, C. N. (2009). Resources, not kinship, determine social patterning in the territorial gunnison's prairie dog (Cynomys gunnisoni). *Ethology, 115*(1), 59–69. doi:10.1111/j.1439-0310.2008.01590.x

Welch, W. J. (1990). Construction of permutation tests. *Journal of the American Statistical Association, 85*, 693–698. doi:10.1080/01621459.1990.10474929

Whitehead, H. (1999). Testing association patterns of social animals. *Animal Behaviour, 57*, F26–F29. doi:10.1006/anbe.1999.1099

Whitehead, H. (2008). Precision and power in the analysis of social structure using associations. *Animal Behaviour*, *75*, 1093–1099. doi:10.1016/j.anbehav.2007.08.022

Whitehead, H. (2009). SOCPROG programs: analysing animal social structures. *Behavioral Ecology and Sociobiology, 63*(5), 765–778. doi:10.1007/s00265-008-0697-y

Williams, N. M., & Thomson, J. D. (1998). Trapline foraging by bumble bees: III. Temporal patterns of visitation and foraging success at single plants. *Behavioral Ecology*, *9*(6), 612–621. doi:10.1093/beheco/9.6.612

Woodruff, P. W. R., Wright, I. C., Shuriquie, N., Russouw, H., Rushe, T., Howard, R. J., Graves, M., et al. (1997). Structural brain abnormalities in male schizophrenics reflect fronto-temporal dissociation. *Psychological Medicine*, *27*(06), 1257–1266. doi:10.1017/S0033291797005229

Wright, D., Butlin, R. K., & Carlborg, Ö. (2006). Epistatic regulation of behavioural and morphological traits in the zebrafish (Danio rerio). *Behavior Genetics*, *36*(6), 914–922. doi:10.1007/s10519-006-9080-9

Yang, W., Stokes, G. S., & Hui, C. H. (2005). Cross-cultural validation of Holland's interest structure in Chinse population. *Journal of Vocational Behavior*, *67*(3), 379–396. doi:10.1016/j.jvb.2004.08.003

Yoder, P. J., Blackford, J. U., Waller, N. G., & Kim, G. (2004). Enhancing power while controlling family-wise error: An illustration of the issues using electrocortical studies. *Journal of Clinical and Experimental Neuropsychology*, *26*(3), 320–331. doi:10.1080/13803390490510040

Yoder, P. J., Bruce, P., & Tapp, J. (2001). Comparing sequential associations within a single dyad. *Behavior Research Methods, Instruments, & Computers*, *33*(3), 331–338. doi:10.3758/BF03195386

Yue, W., Yang, Y., Zhang, Y., Lu, T., Hu, X., Wang, L., Ruan, Y., et al. (2011). A case-control association study of NRXN1 polymorphisms with schizophrenia in Chinese Han population. *Behavioral and Brain Functions*, *7*(1), 7. doi:10.1186/1744-9081-7-7

Zhang, F., St Clair, D., Liu, X., Sun, X., Sham, P. C., Crombie, C., Ma, X., et al. (2005). Association analysis of the RGS4 gene in Han Chinese and Scottish populations with schizophrenia. *Genes, Brain and Behavior*, *4*(7), 444–448. doi:10.1111/j.1601-183X.2005.00167.x

Zhao, J., Quyyumi, A. A., Patel, R., Zafari, A. M., Veledar, E., Onufrak, S., Shallenberger, L. H., et al. (2009). Sex-specific association of depression and a haplotype in leukotriene A4 hydrolase gene. *Psychosomatic Medicine*, *71*(7), 691–696. doi:10.1097/PSY.0b013e3181b05c57

Zheng, F., Wang, L., Jia, M., Yue, W., Ruan, Y., Lu, T., Liu, J., et al. (2011). Evidence for association between Disrupted-in-schizophrenia 1 (DISC1) gene polymorphisms and autism in Chinese Han population: A family-based association study. *Behavioral and Brain Functions*, *7*(1), 14. doi:10.1186/1744-9081-7-14

Zumbo, B. D. (1996). Randomization test for coupled data. *Attention, Perception, & Psychophysics*, *58*(3), 471–478. doi:10.3758/BF03206822

Table 1

Categories of Theoretical Articles, Number of Articles for Each Category, and Articles in Each Category

Category and number of articles for each category	Articles in each category
Introduction to and instruction of PTs (13)	Bai and Pan (2008), Barbella, Denby, and Landwehr (1990), Bear (1995), Chin, Haughton, and Aczel (1996), Hunter and May (2003), LaFleur and Greevy (2009), Lock and Lock (2008), Maguire and Rogers (1989), May and Hunter (1993), Rodgers (1999), Rossman (2008), Tintle, VanderStoep, Holmes, Quisenberry, and Swanson (2011), and Utts (1989)
Algorithms, programs and software for PTs (54)	Bakeman, McArthur, and Quera (1996), Bakeman, Robinson, and Quera (1996), Bejder, Fletcher, and Brager (1998), Berry, Johnston, and Mielke (2005, 2008a), Berry, Johnston, Zahran, and Mielke (2009), Berry and Mielke (1992a, 1992b, 1994, 1996, 1997, 1998), Bilker, Brensinger, and Gur (2004), Buja and Eyuboglu (1992), Bulté and Onghena (2008, 2009), Cai (2006), Chen and Dunlap (1993), de Vries (1993,1995), Dekle, Leung, and Zhu (2008), Dixon, Woodard, and Merry (1998), Dunlap, Burke, and Smith-Crowe (2003), Edgington and Khuller (1992), Eudey, Kerr, and Trumbo (2010), Fischhoff, Dushoff, Sundaresan, Cordingley, and Rubenstein (2009), Gondan (2010), Hayes (1998), Hemelrijk (1990), Johnston, Berry, and Mielke (2008), Lapointe and Garland (2001), Legendre and Lapointe (1995), Linting, van Os, and Meulman (2011), Long, Berry, and Mielke (2009, 2010), May, Masson, and Hunter (1989), McKenzie, Mackinnon, David (1997), Medland, Schmitt, Webb, Kuo, and Neale (2009), Mielke and Berry (1994, 2000), Mundry (1999), Mundry and Sommer (2007), Neuhäuser and Manly (2004), Neuhäuser and Ruxton (2009), O'Connor (1999, 2006), Onghena and May (1995), Rounds, Tracey, and Hubert (1992), Stahl (2004), Tan, Imbos, and Does (1994), Sundaresan, Fischhoff, and Dushoff (2009), Tracey (1997), Vera, Heiser, and Murillo (2007), and Whitehead (2009)
Performance of PTs (27)	Adams and Anthony (1996), Beasley, Erickson, and Allison (2009), Berry, Johnston, and Mielke (2003, 2008b, 2011), Berry, Mielke, and Mielke (2002), Camilli and Smith (1990), Campbell, Legendre, Lapointe (2009), Dekker, Krackhardt, and Snijders (2007), Dunlap (2003), Groppe, Urbach, and Kutas (2011), Hayes (1996), Hayes and Cai (2007), Kimberg,

	Coslett, and Schwartz (2007), Krause et al. (2009), Lapointe and Legendre (1995), Lunneborg (2001), Medina, Kimberg, Chatterjee, and Coslett (2010), Mewhort (2005), Mewhort, Kelly, and Johns (2009), Mielke and Berry (2002), Myors (1998), Oden and Sokal (1992), Peres-Neto and Olden (2001), van den Brink and van den Brink (1989), Whitehead (2008), and Zumbo (1996)
Multivariate PTs (11)	Abbate, Giogianni, Munao, Pesarin, and Salmaso (2001), Blair, Higgins, Karniski, and Kromrey (1994), Blair and Karniski (1993), Finch and Davenport (2009), Maris (2004), Mielke and Berry (1999, 2007a), Mielke, Berry, and Neidt (1996), Pesarin (1990), Yoder, Blackford, Waller, and Kim (2004), Yoder, Bruce, and Tapp (2001)
PTs for single-case designs (22)	Edgington (1996), Ferron and Foster-Johnson (1998), Ferron, Foster-Johnson, and Kromrey (2003), Ferron and Jones (2006), Ferron and Onghena (1996), Ferron and Sentovich (2002), Ferron and Ware (1994, 1995), Fisch (2001), Haardörfer and Gagné (2010), Kratochwill and Levin (2010), Lall and Levin (2004), Levin and Wampold (1999), Levin, Lall, and Kratochwill (2011), Manolov and Solanas (2008, 2009), Manolov, Solanas, Bulté, and Onghena (2010), Onghena (1992), Onghena and Edgington (1994), Perdices and Tate (2009), Sierra, Solanas, and Quera (2005), and Solanas, Sierra, Quera, and Manolov (2008)
Advanced topics (14)	Dzhafarov, Ünlü, Trendtel, and Colonius (2011), Johnston, Berry, and Mielke (2007a, 2007b), Lix and Sajobi (2010), Long, Berry, and Mielke (2007), McKenzie, Onghena, Hogenraad, Martindale, and MacKinnon (1999), Mewhort, Johns, and Kelly (2010), Mielke, Berry, and Johnston (2004), Mielke, Johnston, and Berry (2004), Prelog, Berry, and Mielke (2009), Takane, Hwang, and Abdi (2008), Takane and Hwang (2002, 2005), and Verboon (1993)

Table 2

Number and Percentage of Theoretical Articles by Research Categories

Research categories	**Number**	**Percentage**
Introduction to and instruction of PTs	13	9.2
Algorithms, programs and software for PTs	54	38.3
Performance of PTs	27	19.2
Multivariate PTs	11	7.8
PTs for single-case designs	22	15.6
Advanced topics	14	9.9
Total	141	

Table 3

Classification of Application Articles: Active Areas of Application and Types of Analysis

Areas of application/type of analysis	**Article**
Vocational interests (19 articles in total)	
Order relations (19)	Anderson, Tracey, and Rounds (1997), Darcy and Tracey (2007), Day, Rounds, and Swaney (1998), du Toit and de Bruin (2002), Einarsdóttir (2002), Fouad (1997), Glideen-Tracey and Greenwood (1997), Glidden-Tracey and Parraga (1996), Hedrih (2007), Hofsess and Tracey (2005), Leong, Austin, Sekaran, and Komarraju (1998), Long and Tracey (2006), Nagy, Trautwein, and Lüdtke (2010), Rounds and Tracey (1996), Ryan, Tracey, and Rounds (1996), Šverko (2008), Tracey and Rounds (1994), Tracey and Ward (1998), and Yang, Stokes, and Hui (2005)
Animal behavior (24 articles in total)	
Randomness analysis (19)	Avilés, Varas, and Dyreson (1999), de Vries, Steven, and Vervaecke (2006), Durrell, Sneddon, O'Connell, and Whitehead (2004), Franks, Sendova-Franks, and Anderson (2001), Fujita and Higuchi (2007), Kennedy, Spencer, and Gray (1996), Lemons and Sedinger (2011), Miller, Shapiro, Tyack, and Solow (2004), Mitani and Amsler (2003), Ramos-Fernandez et al. (2009), Ramp, Hagen, Palsbøll, Bérubé, and Sears (2010), Ribi and Porter (1995), Richardson, Ewen, Amstrong, and Hauber (2010), Roberts and Evans (1993), Sendova-Franks and Franks (1995a, 1995b), Trainer, McDonald, and Learn (2002), Verdolin and Slobodchikoff (2009), and Williams and Thomson (1998)
Single-case analysis (1)	Koene (1996)
Proximity matrices (3)	Mitani and Brandt (1994), Mitani, Merriwether, and Zhang (2000), and Payne, Woods, Siddall, and Parr (2000)
Multivariate analysis (1)	Normando, Corain, Salvadoretti, Meers, and Valsecchi (2009)

Brain and behavior (9 articles in total)	
Multiple comparisons (4)	Apostolova et al. (2008), Chakrabarti, Bullmore, and Baron-Cohen (2006), Chua et al. (2009), and Dineen et al. (2009)
Multivariate analysis (1)	Kayser, Tenke, Gates, and Bruder (2007)
Repeated-measure analysis (1)	Knyazeva et al. (2010)
Regression & correlation (2)	Ng, Cooper, Williams, Bullmore, de Zubicaray, and Suckling (2001) and Ridler et al. (2001)
Independent two-group comparison (1)	Woodruff et al. (1997)
Gene and behavior (12 articles in total)	
Multiple comparisons (11)	Arias-Vasquez et al. (2011), Chakrabarti et al. (2009), Hahn et al. (2008), Lerer et al. (2010), Liao et al. (2009), Savitz, van der Merwe, and Ramesar. (2008), Wright, Butlin, and Carlborg (2006), Yue et al. (2011), Zhang et al. (2005), Zhao et al. (2009), and Zheng et al. (2011)
Regression & correlation (1)	Ferraro et al. (2011)
Educational psychology (7 articles in total)	
Single-case analysis (5)	Bardon, Dona, and Symons (2008), Dennin and Ellis (2003), Huang and Chao (1999), Regan, Mastropieri, and Scruggs (2005), and Ross and Begeny (2011)
Randomness analysis (2)	Goldstein, Arnold, Rosenberg, Stowe, and Ortiz (2001) and Kessler and Treiman (2001)
Neuropsychology (6 articles in total)	
Proximity matrices (1)	Brusco (2004)
Regression & correlation (1)	Carlier, Beau, Marchaland, and Michel (1994)
Multivariate analysis (1)	Fehr, Wiedenmann, and Herrmann (2007)
Independent two-group comparison (1)	Longenecker, Kohn, Liu, Zoltick, Weinberger, and Elvevag (2010)
Single-case analysis (2)	Samuel et al. (2000) and Sunderland, Walker, and Walker (2006)
Multidisciplinary psychology (3 articles in total)	

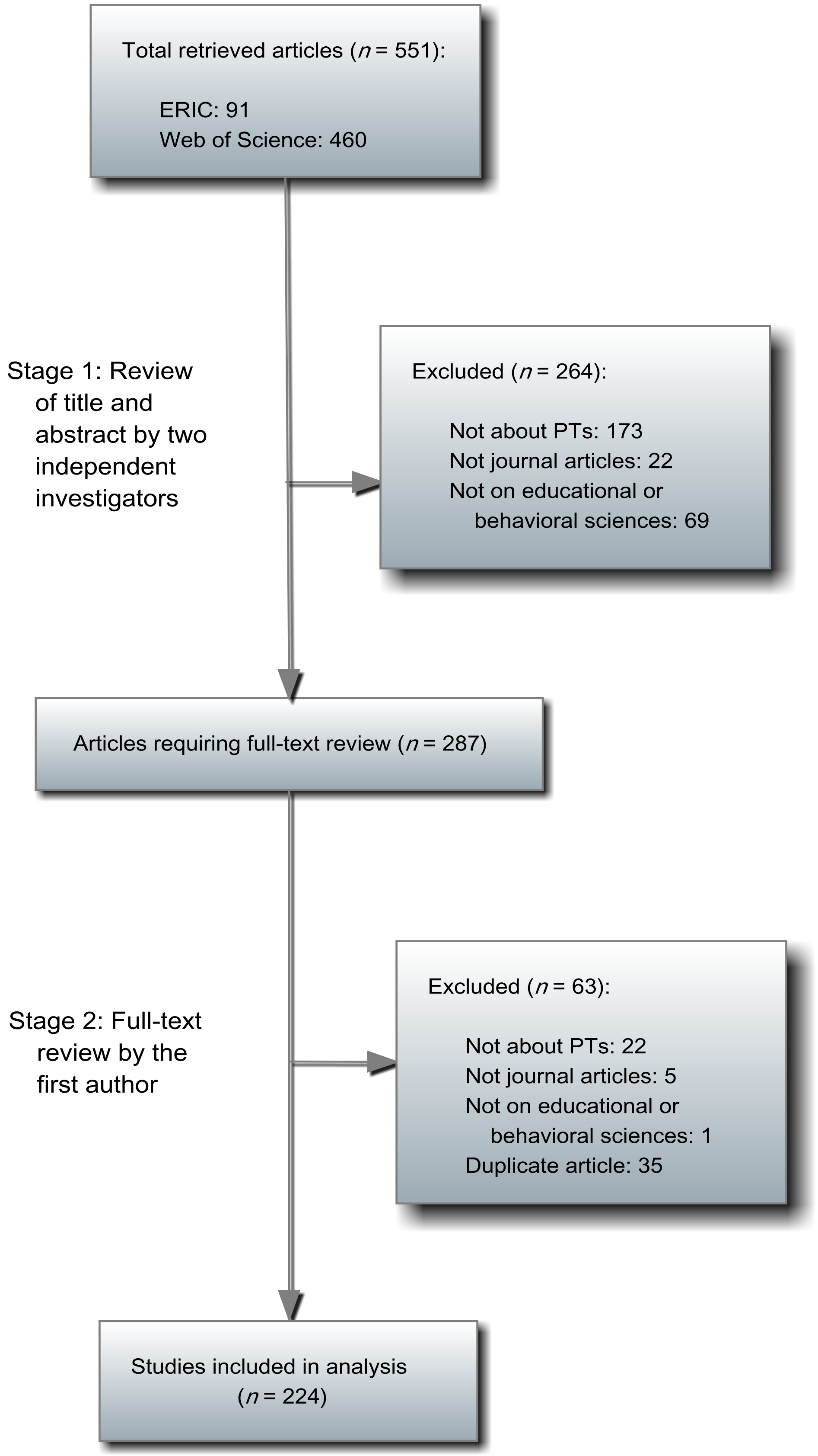
Independent two-group analysis (2)	Greeff and Conradie (1998) and Spilková and Hochel (2009)
Randomness analysis (1)	Gutierrez-Garcia and Tusell (1997)
Education research (3 articles in total)	
Paired two-group comparison (2)	Boltri, Hash, and Vogel (2003) and Snyder et al. (2010)
Independent two-group comparison (1)	Kline (1991)

Table 4

Number and Percentage of Application Articles by Relevant Types of Analysis

Statistical methods	**Number**	**Percentage**
Independent two-group comparison	5	6.2
Paired two-group comparison	2	2.5
Regression & correlation	4	4.9
Repeated-measures analysis	1	1.2
Multivariate analysis	3	3.7
Multiple comparisons	15	17.3
Single-case analysis	8	9.9
Proximity matrices	4	4.9
Order relations	19	23.5
Randomness analysis	22	25.9
Total	83	

Figure 1. Study flow diagram.



Appendix: Search Strategy

ERIC Search

“randomization test*” OR “randomisation test*” OR “permutation test*”

Limits: English, journal articles, timespan (from 1989 to 2011)

Total number of citations: 91

Web of Science Search

Stage 1: Topic = (randomization test*) or Topic = (randomisation test*) or Topic = (permutation test*) Refined by: Document Type = (ARTICLE) and Language = (ENGLISH) Databases = SCI-EXPANDED, SSCI, A&HCI Timespan = (from 1989 to 2011)

Total number of citations: 7281

Stage 2: Afterwards, we refined the total 7281 articles by: Web of Science Categories = (Mathematics interdisciplinary applications) OR (Psychology Experimental) OR (Behavioral sciences) OR (Psychology multidisciplinary) OR (Psychology mathematical) OR (Psychology applied) OR (Psychology developmental) OR (Psychology) OR (Social sciences mathematical methods) OR (Psychology clinical) OR (Multidisciplinary sciences) OR (Psychology educational) OR (Education educational research)

Obtained number of citations after the second stage: 460