

## Useful experimental designs and rank order statistics in educational research

Zendler, Andreas ✉

University of Education Ludwigsburg, Germany ([zendler@ph-ludwigsburg.de](mailto:zendler@ph-ludwigsburg.de))

Vogel, Markus

University of Education Heidelberg, Germany ([vogel@ph-heidelberg.de](mailto:vogel@ph-heidelberg.de))

Spannagel, Christian

University of Education Heidelberg, Germany ([spammagel@ph-heidelberg.de](mailto:spammagel@ph-heidelberg.de))



ISSN: 2243-7703  
Online ISSN: 2243-7711

OPEN ACCESS

**Received:** 1 July 2012

**Revised:** 10 August 2012

**Accepted:** 1 September 2012

**Available Online:** 8 September 2012

**DOI:** 10.5861/ijrse.2012.145

### ***Abstract***

Experimental educational research is of great impact because it illuminates cause-and-effect relationships by accumulating empirical evidence. The present article does not propose new methods but brings three useful experimental designs as well as appropriate statistical procedures (rank order statistics) to the attention of the reader to conduct educational experiments, even with small samples. By means of their systematic use combined with the process-product paradigm of experimental educational research, the influence of essential variables (teacher, context, and process variables) in schools, universities, and other educational institutions can be investigated. The statistical procedures described in this article guarantee that small samples (e.g. a school class) can be successfully used, and that product variables (e.g. knowledge, comprehension, transfer) are only required to meet the criteria of an ordinal scale. The experimental designs and statistical procedures are exemplified by hypothetical data and detailed calculations.

***Keywords:*** experimental education; experimental design; non-parametric statistics; research methodology; statistics

## Useful experimental designs and rank order statistics in educational research

### 1. Introduction

The experiment "has been one of science's most powerful methods for discovering descriptive causal relationships" (Shadish, Cook, & Campbell, 2002, p. 26). A close look at journals of experimental educational research (International Review of Education; Journal of Education and Behavioral Statistics; Journal of Experimental Education; American Educational Research Journal; Review of Educational Research; British Educational Research Journal; Zeitschrift für empirische Pädagogik, Empirische Pädagogik) from 1998 through mid of 2012, however, leads to the conclusion that besides numerous reports on empirical studies only few articles exist dealing with methodological issues of experimental educational research:

- Research on applied parametric and non-parametric statistics (Vargha & Delaney, 1998; Zimmermann, 1998; Luh, 1999; MacDonald, 1999, Mielke & Berry, 1999; Beasley, 2000; Luh & Guo, 2006)
- Research on statistical significance and effect size (MacDonald, 1999; DeVaney, 2001; Thompson & Snyder, 1997; Thompson, 2001; Wilcox, 2006; Hedges, 2007; Schochet, 2008; Raykov & Marcoulides, 2010; Hedges 2011; Aloe & Becker, 2012)
- Research on drop-outs, on handling missing values, and on data re-scaling (Harwell & Gatti, 2001; Schwippert, 2002; Zhang & Rubin, 2003; Moerbeek, 2008; Barnard & Rubin 2010; Shin & Raudenbush, 2010; Pietsch, 2011)

These articles highlight important aspects relevant to experimental educational research. However, the following important issues are underrepresented in the journals cited above:

- Experiments must be amenable to being carried out with small samples (e.g. size of a class),
- Measures of learning are often based on an ordinal scale type and must therefore be analyzed by means of appropriate statistics (cited from Zendler & Pfeiffer, 2009).

Based on these considerations, the present article aims at providing useful factorial designs as well as appropriate statistical procedures (*rank order statistics*) for experimental educational research. Thereby, we are reacting to Kieffer, Reese, and Thompson (2001), who expressed the following requirement in conjunction with the methodological improvement of empirical educational research: "*Those of us who are methodologists bear a responsibility to do our best to move the field toward improved practice. Our colleagues expect us to offer leadership, because given knowledge explosions within their disciplines; they have difficulty enough remaining current in their own areas, much less with emerging developments in statistics techniques and thinking*" (Kieffer, Reese, & Thompson, 2001, p. 305).

The present article consists of five sections: Section 2 gives an overview of the scope of educational research. The third section deals with methodological demands on educational experiments. Section 4 introduces three experimental designs and statistical procedures (*rank order statistics*) which are adequate for experimental education research. Conclusions are drawn in the last section. The appendices illustrate the principle of *rank order statistics* and provide calculations for the examples described in section 4.

### 2. The scope of educational research

Educational research is the scientific investigation of education, its determining variables, its internal processes and the results of these processes (cited from Shuell, 1996; Helmke, 2005, 2006). Modern educational research (Shuell, 1996) is based on the process-product paradigm originally developed by Dunkin and Biddle

(1974). This paradigm takes into account four main factors relevant to learning and teaching: (1) presage variables such as teacher characteristics, type of teacher preparation program, etc; (2) context variables such as student characteristics, and characteristics of the school and classroom environment; (3) process variables such as actual activities of classroom teaching and learning; and (4) product variables such as outcomes of teaching, i.e. what the students have learned. Thus, the process-product paradigm focuses on how processes in the classroom affect products, e.g. learning outcomes. Enhancements of the process-product paradigm address basic conditions of teaching, classroom environments, and mediation processes (cited from Shuell, 1996; Wellenreuther, 2007).

*Basic conditions of teaching.* Basic conditions of teaching deal with two aspects: on the one hand characteristics of the class which are unmodifiable by the teacher prior to instruction; on the other hand factors such as the catchment area of the school and the composition of the class (e.g. percentage of students with foreign mother tongue, percentage of students from classes of lower income, level of previous knowledge at the beginning of instruction).

*Classroom environments.* According to Doyle (1986) this factor has six properties related with the dynamics of classroom instructions: (i) multidimensionality (classrooms are crowded settings with a large number of interconnected events), (ii) simultaneity (many things happen at the same time), (iii) immediacy (a large number of events progress at a rapid rate), (iv) unpredictability (it is difficult how a particular activity will turn out on any given day), (v) publicness (classrooms are public places, there are few places for either the teacher or students to hide), (vi) history (common history of experiences, norms and routines accumulate).

*Mediation processes.* Teacher activities do not directly influence the cognitive, affective, or motivational processes of the students. Moreover, individual student processes mediate the teacher's activities. Negatively expressed: "It is conceivable that a student would fail to learn from a particular instructional setting, regardless of its 'quality', because of an inability to recognize and interpret cues that signal which performances are being taught" (Doyle, 1986, p. 185).

### **3. Methodological demands on experimental educational research**

Experimental educational research may be understood as the scientific sub-discipline of empirical educational research that is engaged in the experimental investigation of teaching and learning in the context of educational institutions. It comes hand in hand with great demands on experimental design and statistical data analysis.

#### *3.1 Experimental design*

*Involvement of several factors.* Several factors (presage, context, process, product variables) should be included in the experimental design in terms of the process-product paradigm. The inclusion of several factors is interesting from a pedagogical point of view because interactions may be tested. For example, in the case that the interaction between method and student characteristics is significant (method  $\times$  student characteristics), different methods have different effects on the dependent variable (e.g. learning outcome) with respect to different learner characteristics (cf. Cronbach & Snow, 1977; O'Malley, 1988; Marjoribanks, 1988).

*Consideration of classroom environment.* The factor *classroom environment* should be explicitly considered in the experimental design. By including classroom environment in a two or higher order factorial design, for instance, it can be tested whether a method concerning different classes – *method  $\times$  class* – has an effect on teaching and learning (cited from Klauer, 1973; 2005). If the interaction is significant, evidence is provided that the method has different effects regarding different classes. In such a case, the hypothesis must be abandoned that a method is generally superior. Furthermore, this demand can also mean that different characteristics of classes (e.g. knowledge level of a school class, students of different social background) may be considered in experimental educational research. Thus, interactions such as *method  $\times$  characteristic of school class* are

testable.

*Usage of an economical design.* The preparation of an economical experimental design depends on the background knowledge of the researcher, the involved hypotheses, the sample size, and the number of levels of factor combinations. To provide evidence of an experimental effect without generalizations to the population, relative small samples are sufficient. The demand on economic experimental design in the context of educational experiments is discussed in the literature (Lewis, 1968; Cooper, 1988; Klauer, 1973; 2005).

### 3.2 Statistical analysis

Keselman et al. (1998) investigate the uses of data analysis in 226 articles of several prominent educational journals (e.g. American Educational Research Journal, Journal of Educational Computing Research, and Journal of Experimental Education). They state: "*In particular, applied statisticians have devoted a great deal of effort to understanding the operating characteristics of statistical procedures when the distributional assumptions that underlie a particular procedure are not likely to be satisfied. It is common knowledge that, under certain data-analytic conditions, statistical procedures will not produce valid results. The applied researcher who routinely adopts a traditional procedure without giving thought to its associated assumptions may unwittingly be filling the literature with nonreplicable results*" (Keselman, 1998, p. 351). Besides the assumptions that Keselman et al. stress, *the scale type of product variables* and *the sample size* are of utmost importance to consider when conducting educational experiments.

*Assumptions.* From many methodological research reviews (Goodwin & Goodwin, 1985; West, Carmody, & Stallings, Keselman et al., 1998; Authors, 2009) it is known that parametric tests (*t*-test, *F*-test, etc.) are used to analyze educational experiments without taking into account their distributional assumptions (normal distribution, independently distributed homogeneity of variance). Non-parametric statistics are an interesting alternative; most frequently they simply require that the population is continuous (Gibbons & Chakraborti, 2003). Although the assumptions are lower than for parametric tests, it is certainly true that variance homogeneity is a prerequisite for some non-parametric statistics, e.g. the Kruskal-Wallis test (cited from Keselman, 1998; Lix et al., 1996).

*Scale type of product variables.* Data from experimental educational research (e.g. knowledge, comprehension, transfer data) must be considered ordinal unless the properties of an interval scale have been verified from a measurement theoretical point of view (Krantz, Luce, Suppes, & Tversky, 1971; Roberts, 1979; Narens, 2007). Thus, *rank order statistics* should be used for data analysis (Cooper, 1988; Gibbons & Chakraborti, 2003; Kvam & Vidakovic, 2007).

*Sample size (N = small).* Due to the involvement of classroom environments in experimental educational research, normally small samples must be used. For this, statistical tests should be applied which consider this restriction. *Rank order statistics* are appropriate in cases with small samples (Lienert, 1978; Hettmansperger, 1984; Siegel & Castellan, 1988, Cooper, 1988). There is no simple answer to the question what *small* means (Huck, 2009). Some authors recommend rank order statistics to be used when sample size is as small as 5 or 6 subjects per group (Siegel, 1956; Pett, 1997; Huck, 2009).

## 4. Experimental designs and appropriate rank order statistics

Here we can see that three experimental designs are described whose necessity is a result of demands on experimental educational research (see section 3), empirical findings (Authors, 2009; Wellenreuther, 2007) and theoretical requirements (Klauer, 1973; 2005). In detail, the following three designs are explicated using the nomenclature of Kirk (1995):

- CRF- $pq$  experimental design<sup>1</sup>
- SPF- $p•q$  experimental design<sup>2</sup>
- SPF- $pr•q$  experimental design<sup>3</sup>

For each of the three experimental designs *rank order statistics* are introduced and exemplified by data sets. These data sets are hypothetical because no educational experiments have been published applying these statistical procedures. It should be noted that standard commercial software (SPSS, SAS) does not support advanced *rank order statistics*, but they can be easily implemented using spreadsheet calculation programs.

There are no effect size indices reported for the statistical procedures presented in this article. Nevertheless, the effect size indicator  $r_{equivalent}$  introduced by Rosenthal and Rubin (2003) can be used.  $r_{equivalent}$  is an effect size estimate for statistical procedures where no generally accepted effect size indicator exists.

In this article only factorial designs are considered in which all possible combinations of factor levels occur. Especially when whole classes participate in a study, hierarchical designs can be used. In this case, the hierarchical linear model (HLM) is appropriate for statistical analyses (Bryk & Raudenbush, 1992; Hox, 2000), even with ordinal data (Raudenbush & Bryk, 2002; Hedeker 2008). Although the HLM truly partitions variation among levels, it has more restrictive assumptions than rank order statistics presented here. For example, a sufficient large number of units on each level are necessary: for a two-level analysis conducted in schools, a large number of schools and students must participate (Mok, 1995).

#### 4.1 CRF- $pq$ experimental designs to test interactions

CRF- $pq$  experimental designs allow for investigating the effects of two factors  $A$  and  $B$  on a dependent variable. The factor  $A$  may have  $p$ , the factor  $B$  may have  $q$  levels.

CRF- $pq$  experimental designs are important for experimental educational research because they allow for investigating two main effects, as well as the investigation of pedagogically important interaction effects such as *method*  $\times$  *media support* or *method*  $\times$  *characteristics of media support*.

The characteristics of CRF- $pq$  experimental designs are as follows: (1)  $p \geq 2$  levels of learning and teaching (e.g.  $p$  different methods) may be considered for factor  $A$ ; (2)  $q \geq 2$  kinds of media support (specific characteristics) may be considered for factor  $B$ ; (3) Learners are randomly assigned to the levels of factor combinations; (4)  $pq$  samples of learners are required. If each sample consists of  $n$  learners, then the total sample size is  $N=pqn$ .

Figure 1 depicts the layout of a CRF-32 design with  $p = 3$  levels of teaching ( $a_1, a_2, a_3$ ) and learning and  $q = 2$  levels of the factor classroom (resp. classroom characteristics or learner characteristics) ( $b_1, b_2$ ) resulting in six samples  $S_1$  through  $S_6$ .

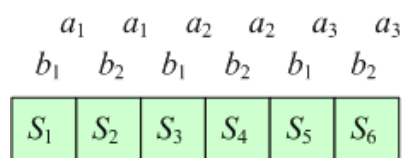


Figure 1. Layout of a CRF-32 experimental design

<sup>1</sup> CRF = Completely Randomized Factorial;  $p$  and  $q$  denote the number of levels of the two factors

<sup>2</sup> SPF = Split-Plot Factorial;  $p$  and  $q$  denote the number of levels of the two factors

<sup>3</sup> SPF = Split-Plot Factorial;  $p$ ,  $r$ , and  $q$  denote the number of levels of the three factors

**Rank order statistics for CRF-pq designs: The bifactorial H-test**

The bifactorial  $H$ -test by Lemmer and Stoker (1967) allows for testing the effects of a factor  $A$  with  $p$  levels and a factor  $B$  with  $q$  levels on a dependent variable. Equal samples sizes per cell are assumed. In addition, subjects are completely randomized regarding factor  $A$  and factor  $B$ .

1. *Statistical hypotheses:* The bifactorial  $H$ -test allows for testing three null hypotheses regarding the factors  $A$  and  $B$  as well as the interaction  $A \times B$ . The hypotheses are:

i) the means  $\mu_1, \mu_2, \dots, \mu_p$  of the populations with the  $p$  levels of factor  $A$  are equal, i.e.

$$H_0^A : \mu_1 = \mu_2 = \dots = \mu_p;$$

ii) the means  $\mu_1, \mu_2, \dots, \mu_q$  of the populations with the  $q$  levels of factor  $B$  are equal, i.e.

$$H_0^B : \mu_1 = \mu_2 = \dots = \mu_q;$$

iii) the means  $\mu_{11}, \mu_{12}, \dots, \mu_{pq}$  of the populations with the  $pq$  levels of the factor combination  $A \times B$  are equal, i.e.

$$H_0^{A \times B} : \mu_{11} = \mu_{12} = \dots = \mu_{pq}.$$

2. *Test statistics:* The test statistics of the bifactorial  $H$ -test are based on three independent components  $H_A$ ,  $H_B$ , and  $H_{A \times B}$  connected by addition (cited from Lemmer & Stoker, 1967, p. 70):

$$H = H_A + H_B + H_{A \times B}.$$

The test statistics are formulated in analogy to the classical  $H$ -test by Kruskal and Wallis (1952). For this, the overall statistics  $H = \frac{12}{N \cdot (N + 1)} \cdot \sum_{i=1}^{p \cdot q} \frac{T_i^2}{n_i} - 3 \cdot (N + 1)$  is used;  $n_i$  are sample sizes for the individual levels,  $N = \sum n_i$  is the total sample size;  $T_i$  are rank sums of the individual samples.

$$\text{for } H_0^A : H_A = \frac{12}{N \cdot (N + 1)} \cdot \sum_{i(a)=1}^p \frac{T_{i(a)}^2}{n_{i(a)}} - 3 \cdot (N + 1);$$

$$\text{for } H_0^B : H_B = \frac{12}{N \cdot (N + 1)} \cdot \sum_{i(b)=1}^q \frac{T_{i(b)}^2}{n_{i(b)}} - 3 \cdot (N + 1);$$

$$\text{for } H_0^{A \times B} : H_{A \times B} = H - H_A - H_B.$$

The statistics  $H_A$  has a  $\chi^2$  distribution with  $df = p - 1$  degrees of freedom,  $H_B$  has a  $\chi^2$  distribution with  $df = q - 1$  and  $H_{A \times B}$  has a  $\chi^2$  distribution with  $df = (p - 1)(q - 1)$  degrees of freedom.

3. *Decisions:* For three simultaneous  $H$ -tests the decisions at adjusted  $\alpha^* = \alpha/3$  are as follows<sup>4</sup>: (i)  $H_0^A$  is

---

<sup>4</sup> We use the Bonferroni procedure for  $\alpha$ -level adjustment to simplify calculations. In studies with real data more appropriate methods for adjusting the  $\alpha$ -level can be used, such as the Holm procedures (Holm, 1979; Shaffer, 1986).

rejected, if  $H_A > \chi^2_{(\alpha^*; p-1)}$ ; (ii)  $H_0^B$  is rejected, if  $H_B > \chi^2_{(\alpha^*; q-1)}$ ; (iii)  $H_0^{A \times B}$  is rejected, if  $H_{A \times B} > \chi^2_{(\alpha^*; (p-1)(q-1))}$ .

4. *Example:* An educational experiment using a CRF-32 design investigates the effect of three educational methods ( $a_1, a_2, a_3$ ) and two kinds of media support ( $b_1, b_2$ ) on learning to read. Table 1 contains the hypothetical data set – ranks and data (*italics*) – for the CRF-32 design (for calculations see Appendix B.1).

**Table 1**

*Hypothetical data set for the CRF-32 experimental design*<sup>5</sup>

	$a_1$	$a_1$	$a_2$	$a_2$	$a_3$	$a_3$
	$b_1$	$b_2$	$b_1$	$b_2$	$b_1$	$b_2$
	9 (6)	1 (1.5)	20 (12)	12 (8)	29 (28.5)	21 (12.5)
	2 (2)	10 (6.5)	18 (11)	11 (7)	30 (29)	24 (14)
	6 (4.5)	3 (2.5)	19 (11.5)	13 (8.5)	27 (27)	23 (13.5)
	4 (3)	7 (5)	17 (10.5)	14 (9)	28 (28)	22 (13)
	5 (4)	8 (5.5)	16 (10)	15 (9.5)	26 (26.5)	25 (16)
<b><math>T_i</math></b>	<b>26</b>	<b>29</b>	<b>90</b>	<b>65</b>	<b>140</b>	<b>115</b>

The statistics is  $H = 27.43$  (for calculations see Appendix B).

For the pooled data regarding factor  $A$  the statistics is  $H_A = 12.50$  [ $0.0018 < p(\chi^2_{(2)}) < 0.002$ ] that is significant at adjusted  $\alpha^* = 0.0166$ : learning to read differs with respect to the three educational methods.

For the pooled data regarding factor  $B$  the statistics is  $H_B = 6.82$  [ $0.0090 < p(\chi^2_{(1)}) < 0.0091$ ], which is significant at adjusted  $\alpha^* = 0.0166$ : learning to read differ with respect to the two kinds of media support.

The test statistics for the interaction is  $H_{A \times B} = 8.11$  [ $0.0172 < p(\chi^2_{(2)}) < 0.0173$ ] that is significant at adjusted  $\alpha^* = 0.0166$ : the three methods have different effects on learning supported by the two kinds of media.

***Further rank order statistics for CR-pq experimental designs***

Wolfe, Dean, and Hartlaub (1990) present a test optimized for interaction effects. Brunner, Puri, and Sun (1995) introduce a test for stratified samples. Akritas, Arnold, and Brunner (1997) describe a test that can be used with unequal sample sizes.

**4.2 SPF-p•q experimental designs to investigate learning efficiency**

SPF-p•q experimental designs allow for investigating the effects of a factor  $A$  and a repeated measures factor  $B$  on a dependent variable. The factor  $A$  may have  $p$ , the factor  $B$  may have  $q$ -levels.

SPF-p•q experimental designs are of utmost importance for experimental educational research because they are well-suited to investigate learning processes with respect to the product-process paradigm (e.g. one presage

<sup>5</sup>  $T$  denotes the rank sum of each sample.

variable, one context variable, one process variable).

The characteristics of SPF- $p \cdot q$  experimental designs are as follows: (1)  $p \geq 2$  levels of learning and teaching may be considered for factor  $A$ ; (2)  $q \geq 2$  repeated measurements may be considered for factor  $B$ ; (3) learners with respect to the individual blocks are treated  $q$ -fold under the same conditions of teaching and learning; (4)  $p$  samples of learners are necessary. If each sample consists of  $n$  learners, then the total sample size is  $N=pn$ .

Figure 2 shows the layout of a SPF-3•2 design with  $p = 3$  methods ( $a_1, a_2, a_3$ ) and  $q = 2$  levels of repeated measurements ( $b_1, b_2$ ) resulting in three samples  $s_1 \dots s_n, s_{n+1} \dots s_{2n}, s_{2n+1} \dots s_{3n}$ .

	$b_1$	$b_2$
$a_1$	$s_1$	$s_1$
	...	...
	$s_n$	$s_n$
$a_2$	$s_{n+1}$	$s_{n+1}$
	...	...
	$s_{2n}$	$s_{2n}$
$a_3$	$s_{2n+1}$	$s_{2n+1}$
	...	...
	$s_{3n}$	$s_{3n}$

Figure 2. Layout of a SPF-3•2 experimental design

**Rank order statistics for SPF- $p \cdot q$  experimental design: The bifactorial  $D^{p \cdot q}$ -test**

The bifactorial  $D^{p \cdot q}$ -test published by Bredenkamp (1974) allows for testing the effects of a factor  $A$  with  $p$  levels and a repeated measures factor  $B$  with  $q$  levels on a dependent variable. Equal samples sizes per cell are assumed. In addition, subjects are completely randomized regarding factor  $A$ .

1. *Statistical hypotheses:* The bifactorial  $D^{p \cdot q}$ -test allows for testing two null hypotheses with respect to the factor  $B$  and the interaction  $A \times B$ <sup>6</sup>:

i) the means  $\mu_1, \mu_2, \dots, \mu_q$  of the populations under  $q$  levels of factor  $B$  are equal, i.e.

$$H_0^B : \mu_1 = \mu_2 = \dots = \mu_q;$$

ii) the means  $\mu_{11}, \mu_{12}, \dots, \mu_{pq}$  of the populations under  $pq$  levels of the factor combination  $A \times B$  are equal, i.e.

$$H_0^{A \times B} : \mu_{11} = \mu_{12} = \dots = \mu_{pq}.$$

2. *Test statistics:* The test statistics of the bifactorial  $D^{p \cdot q}$ -test is based on the H-test by Kruskal-Wallis. For

this,  $D_{B|A}^{p \cdot q} = \frac{12}{n_q \cdot q \cdot (q+1)} \sum_{i=1}^{p \cdot q} T_i^2 - 3 \cdot n_q \cdot p \cdot (q+1)$  is used,  $N$  is the total sample size,  $n_q$  are the sample

sizes for the individual levels of factor  $B|A$ ,  $T_i$  are rank sums of the individual samples.

<sup>6</sup> At this point, it should be mentioned that the hypotheses expressed deal with the means of the distributions. Bruner, Dette, and Munk (1997) discuss in more detail hypotheses in nonparametric factorial designs.



The test statistics are:

$$\text{for } H_0^B: D_B^{p,q} = \frac{12}{N \cdot q \cdot (q+1)} \sum_{i(b)=1}^q T_{i(b)}^2 - 3N \cdot (q+1);$$

$$\text{for } H_0^{A \times B}: D_{A \times B}^{p,q} = D_{B|A}^{p,q} - D_B^{p,q}.$$

The statistics  $D_B^{p,q}$  has a  $\chi^2$  distribution with  $df = q-1$  degrees of freedom,  $D_{A \times B}^{p,q}$  has a  $\chi^2$  distribution with  $df = (p-1)(q-1)$  degrees of freedom.

3. *Decisions:* For two simultaneous  $D^{p,q}$ -tests, the decisions at adjusted  $\alpha^* = \alpha/2$  are as follows: (i)  $H_0^B$  is rejected, if  $D_B^{p,q} > \chi_{(\alpha^*; q-1)}^2$ ; (ii)  $H_0^{A \times B}$  is rejected, if  $D_{A \times B}^{p,q} > \chi_{(\alpha^*; (p-1)(q-1))}^2$ .

4. *Example:* An educational experiment using a SPF-3•2 design investigates the effect of three educational methods ( $a_1, a_2, a_3$ ) and at two different measuring times ( $b_1, b_2$ ) on the comprehension of the Pythagorean Theorem. Table 2 contains the hypothetical data set – ranks and data (*italics*) – for the SPF-3•2 design (for calculations see Appendix B.2).

**Table 2**

*Hypothetical data set for a SPF-3•2 experimental design*

		<b><math>b_1</math></b>	<b><math>b_2</math></b>
<b><math>a_1</math></b>	$s_1$	1 ( <i>17</i> )	2 ( <i>19</i> )
	$s_2$	1 ( <i>14</i> )	2 ( <i>25</i> )
	$s_3$	1 ( <i>13</i> )	2 ( <i>24</i> )
	$s_4$	2 ( <i>15</i> )	1 ( <i>14</i> )
	$s_5$	1 ( <i>11</i> )	2 ( <i>16</i> )
	<b><math>T_i</math></b>	<b>6</b>	<b>9</b>
		<b><math>b_1</math></b>	<b><math>b_2</math></b>
<b><math>a_2</math></b>	$s_6$	1 ( <i>12</i> )	2 ( <i>15</i> )
	$s_7$	2 ( <i>16</i> )	1 ( <i>13</i> )
	$s_8$	2 ( <i>14</i> )	1 ( <i>13</i> )
	$s_9$	1 ( <i>15</i> )	2 ( <i>28</i> )
	$s_{10}$	1 ( <i>19</i> )	2 ( <i>20</i> )
	<b><math>T_i</math></b>	<b>7</b>	<b>8</b>
		<b><math>b_1</math></b>	<b><math>b_2</math></b>
<b><math>a_3</math></b>	$s_{11}$	2 ( <i>24</i> )	1 ( <i>3</i> )
	$s_{12}$	2 ( <i>27</i> )	1 ( <i>6</i> )
	$s_{13}$	2 ( <i>18</i> )	1 ( <i>7</i> )
	$s_{14}$	2 ( <i>29</i> )	1 ( <i>8</i> )
	$s_{15}$	2 ( <i>19</i> )	1 ( <i>12</i> )
	<b><math>T_i</math></b>	<b>10</b>	<b>5</b>

The auxiliary formula calculates from  $D_{B|A}^{p,q} = 7.0$ ).

For the pooled data regarding factor  $B$  of the experiment the test statistics is  $D_B^{p,q} = 0.07$  [ $0.79 < p(\chi_{(1)}^2) < 0.80$ ] that is not significant at adjusted  $\alpha^* = \alpha/2 = 0.025$ : comprehension of the Pythagorean theorem does not differ with respect to the two levels of measuring time. Thus, neither method led to a change in the comprehension of the Pythagorean Theorem.

By using the auxiliary formula  $D_{B|A}^{p,q}$  the test statistics  $D_{A \times B}^{p,q} = 6.93$  [ $0.031 < p(\chi_{(2)}^2) < 0.032$ ] that is not significant at  $\alpha^* = 0.025$ : the comprehension of the Pythagorean theorem does not differ for the three educational methods in interaction with the two different measuring times. Thus, there are no group differences in changing the comprehension of the Pythagorean Theorem.

**Further rank order statistics for SPF-p•q experimental designs**

Rashid (1996) introduced two tests to analyze SPF-p•q experimental designs. Furthermore, he presented post hoc tests for contrasts. In Monte Carlo simulations he provided evidence that his tests have a similar power as analogous parametric tests. Brunner and Dette (1992) as well as Steland (1997) describe tests that admit drop outs.

*4.3 SPF-pr•q experimental designs to investigate learning efficiency with respect to classroom environment*

SPF-pr•q experimental designs allow for investigating the effects of two factors  $A, C$  and a repeated measures factor  $B$  on a dependent variable. The factor  $A$  may have  $p$  levels,  $B$  may have  $q$  levels and  $C$  may have  $r$  levels.

SPF-pr•q designs are the trifactorial enhancements of SPF-p•q designs. They are well-suited to investigate learning processes with respect to two factors (e.g., the first to a presage variable, the second to control the classroom environment).

SPF-pr•q experimental designs are important for educational research for the following reasons: (1)  $p \geq 2$  levels of learning and teaching may be considered for factor  $A$  (e.g. presage variables); (2) further  $r \geq 2$  levels of learning and teaching may be considered for factor  $C$  combined with factor  $A$  (e.g. characteristics of the classroom environment); (3)  $q \geq 2$  repeated measurements may be considered for factor  $B$ ; (4) learners from the individual blocks are faced with same learning and teaching treatment  $q$ -fold; (5)  $pr$  samples of learners are necessary. If each sample consists of  $n$  learners, then the total sample size is  $N=prn$ .

	$b_1$	$b_2$	$b_3$
$a_1c_1$	$s_1$	$s_1$	$s_1$
	...	...	...
	$s_n$	$s_n$	$s_n$
$a_1c_2$	$s_{n+1}$	$s_{n+1}$	$s_{n+1}$
	...	...	...
	$s_{2n}$	$s_{2n}$	$s_{2n}$
$a_2c_1$	$s_{2n+1}$	$s_{2n+1}$	$s_{2n+1}$
	...	...	...
	$s_{3n}$	$s_{3n}$	$s_{3n}$
$a_2c_2$	$s_{3n+1}$	$s_{3n+1}$	$s_{3n+1}$
	...	...	...
	$s_{4n}$	$s_{4n}$	$s_{4n}$

Figure 3. Layout of a SPF-22•3 experimental design

Figure 3 depicts the layout of a SPF-22•3 design (repeated measurements for factor  $B$ ) with  $pr = 2 \times 2 = 4$  samples ( $a_1c_1, a_1c_2, a_2c_1, a_2c_2$ ) and  $q = 3$  levels of factor  $B$  ( $b_1, b_2, b_3$ ) resulting in four samples  $s_1 \dots s_n, s_{n+1} \dots s_{2n}, s_{2n+1} \dots s_{3n}, s_{3n+1} \dots s_{4n}$ .

### Rank order statistics for SPF- $pr \bullet q$ designs: The trifactorial $D^{pr \bullet q}$ -test

The trifactorial  $D^{pr \bullet q}$ -test (Bredenkamp, 1974) allows for testing the effects of a factor  $A$  with  $p$  levels, a repeated measurement factor  $B$  with  $q$  levels, and a factor  $C$  with  $r$  levels on a dependent variable. Equal samples sizes per cell are assumed. In addition, subjects are completely randomized regarding factor  $A$  and factor  $C$ .

1. *Statistical hypotheses:* The trifactorial  $D^{pr \bullet q}$ -test allows for testing seven null hypotheses that relate to the following effects: The unconditional effect of factor  $B$ , the conditional effect of factor  $B$  within factor  $C$ , the conditional effect of factor  $B$  within factor  $A$ , the conditional effect of factor  $B$  within the factor combination  $A \times C$ , and the interactions  $B \times A$ ,  $B \times C$ , and  $A \times B \times C$ :

i) the means  $\mu_1, \mu_2, \dots, \mu_q$  of the populations under  $q$  levels of factor  $B$  are equal, i.e.

$$H_0^B : \mu_1 = \mu_2 = \dots = \mu_q;$$

ii) the conditional means  $\mu_{1|1}, \mu_{1|2}, \dots, \mu_{q|r}$  of the populations under  $q$  levels of factor  $B$  within factor  $C$  are equal, i.e.

$$H_0^{B|C} : \mu_{1|1} = \mu_{1|2} = \dots = \mu_{q|r};$$

iii) the conditional means  $\mu_{1|1}, \mu_{1|2}, \dots, \mu_{q|p}$  of the populations under  $q$  levels of factor  $B$  within factor  $A$  are equal, i.e.

$$H_0^{B|A} : \mu_{1|1} = \mu_{1|2} = \dots = \mu_{q|p};$$

iv) the conditional means  $\mu_{1|11}, \mu_{1|12}, \dots, \mu_{q|pr}$  of the populations under  $q$  levels of factor  $B$  within the factor combination  $A \times C$  are equal, i.e.

$$H_0^{B|A \times C} : \mu_{1|11} = \mu_{1|12} = \dots = \mu_{q|pr};$$

v) the means  $\mu_{11}, \mu_{12}, \dots, \mu_{qp}$  of the populations under  $qp$  levels of factor combination  $B \times A$  are equal, i.e.

$$H_0^{B \times A} : \mu_{11} = \mu_{12} = \dots = \mu_{qp};$$

vi) the means  $\mu_{11}, \mu_{12}, \dots, \mu_{qr}$  of the populations under  $qr$  levels of factor combination  $B \times C$  are equal, i.e.

$$H_0^{B \times C} : \mu_{11} = \mu_{12} = \dots = \mu_{qr};$$

vii) the means  $\mu_{111}, \mu_{112}, \dots, \mu_{pqr}$  of the populations under  $pqr$  levels of the factor combination  $A \times B \times C$  are equal, i.e.

$$H_0^{A \times B \times C} : \mu_{111} = \mu_{112} = \dots = \mu_{pqr}.$$

2. *Test statistics:* The test statistics of the trifactorial  $D^{pr \bullet q}$ -test is based on the H-test by Kruskal-Wallis.  $N$  is

the total size,  $T$  are total rank sums,  $T_{a(i)}$  are the rank sums of the levels of factor A,  $T_{b(i)}$  are the pooled rank sums of the levels of factor B,  $T_i$  are the individual rank sums of the levels of the factors A, B, and C.

The test statistics are:

$$\text{for } H_0^B: \quad D_B^{pr,q} = \frac{12}{N \cdot q \cdot (q+1)} \sum_{i(b)=1}^q T_{i(b)}^2 - 3 \cdot N \cdot (q+1);$$

$$\text{for } H_0^{B|C}: \quad D_{B|C}^{pr,q} = \frac{12 \cdot r}{N \cdot q \cdot (q+1)} \sum_{i(a)=1}^{p-q} T_{i(a)}^2 - 3 \cdot N \cdot (q+1);$$

$$\text{for } H_0^{B|A}: \quad D_{B|A}^{pr,q} = \frac{12 \cdot p}{N \cdot q \cdot (q+1)} \sum_{i(c)=1}^{r-q} T_{i(c)}^2 - 3 \cdot N \cdot (q+1);$$

$$\text{for } H_0^{B|A \times C}: \quad D_{B|A \times C}^{pr,q} = \frac{12 \cdot p \cdot r}{N \cdot q \cdot (q+1)} \sum_i T_i^2 - 3 \cdot N \cdot (q+1);$$

$$\text{for } H_0^{B \times A}: \quad D_{B \times A}^{pr,q} = D_{B|A}^{pr,q} - D_B^{pr,q};$$

$$\text{for } H_0^{B \times C}: \quad D_{B \times C}^{pr,q} = D_{B|C}^{pr,q} - D_B^{pr,q};$$

$$\text{for } H_0^{A \times B \times C}: \quad D_{A \times B \times C}^{pr,q} = D_{B|A \times C}^{pr,q} - D_B^{pr,q} - D_{B \times A}^{pr,q} - D_{B \times C}^{pr,q}.$$

The statistic  $D_B^{pr,q}$  is  $\chi^2$ -distributed with  $df=q-1$  degrees of freedom,  $D_{B|C}^{pr,q}$   $\chi^2$ - distributed with  $df=r(q-1)$ ,  $D_{B|A}^{pr,q}$   $\chi^2$ - distributed with  $df=p(q-1)$ ,  $D_{B \times A}^{pr,q}$   $\chi^2$ - distributed with  $df=pq-p-q+1$ ,  $D_{B|A \times C}^{pr,q}$   $\chi^2$ - distributed with  $df=pr(q-1)$ ,  $D_{B \times C}^{pr,q}$   $\chi^2$ - distributed with  $df=rq-r-q+1$ , and  $D_{A \times B \times C}^{pr,q}$   $\chi^2$ - distributed with  $df=q(pr-p-r+1)+(1-r)(p-1)$ .

3. *Decisions:* For seven simultaneous  $D^{pr,q}$ -tests, the decisions at adjusted  $\alpha^* = \alpha/7$  are as follows: (i)  $H_0^B$  is rejected, if  $D_B^{pr,q} > \chi_{(\alpha^*; q-1)}^2$ ; (ii)  $H_0^{B|C}$  is rejected, if  $D_{B|C}^{pr,q} > \chi_{(\alpha^*; r(q-1))}^2$ ; (iii)  $H_0^{B|A}$  is rejected, if  $D_{B|A}^{pr,q} > \chi_{(\alpha^*; (p(q-1)))}^2$ ; (iv)  $H_0^{B|A \times C}$  is rejected, if  $D_{B|A \times C}^{pr,q} > \chi_{(\alpha^*; (pr(q-1)))}^2$ ; (v)  $H_0^{B \times A}$  is rejected, if  $D_{B \times A}^{pr,q} > \chi_{(\alpha^*; (pq-p-q+1))}^2$ ; (vi)  $H_0^{B \times C}$  is rejected, if  $D_{B \times C}^{pr,q} > \chi_{(\alpha^*; (rq-r-q+1))}^2$ , and (vii)  $H_0^{A \times B \times C}$  is rejected, if  $D_{A \times B \times C}^{pr,q} > \chi_{(\alpha^*; (q(pr-p-r+1)+(1-r)(p-1)))}^2$ .

4. *Example:* An educational experiment using a SPF-22•2 design investigates the effect of two class levels ( $a_1, a_2$ ), two methods ( $c_1, c_2$ ) and three different measuring times ( $b_1, b_2, b_3$ ) on the application of computer science modeling. Table 3 contains the hypothetical data set – ranks and data (*italics*) – for the SPF-22•3 design.

**Table 3**

*Hypothetical data set for a SPF-22•3 experimental design*

		$b_1$	$b_2$	$b_3$
$a_1c_1$	$s_1$	3 ( <i>17</i> )	1 ( <i>10</i> )	2 ( <i>11</i> )
	$s_2$	3 ( <i>20</i> )	2 ( <i>15</i> )	1 ( <i>8</i> )
	$s_3$	3 ( <i>28</i> )	2 ( <i>16</i> )	1 ( <i>9</i> )
	$s_4$	2 ( <i>16</i> )	1 ( <i>10</i> )	3 ( <i>22</i> )
	$s_5$	3 ( <i>24</i> )	2 ( <i>16</i> )	1 ( <i>6</i> )
$T_i$		<b>14</b>	<b>8</b>	<b>8</b>
$a_1c_2$	$s_6$	2 ( <i>13</i> )	1 ( <i>9</i> )	3 ( <i>23</i> )
	$s_7$	1 ( <i>4</i> )	2 ( <i>20</i> )	3 ( <i>24</i> )
	$s_8$	1 ( <i>4</i> )	2 ( <i>19</i> )	3 ( <i>25</i> )
	$s_9$	2 ( <i>14</i> )	1 ( <i>3</i> )	3 ( <i>19</i> )
	$s_{10}$	1 ( <i>14</i> )	2 ( <i>17</i> )	3 ( <i>28</i> )
$T_i$		<b>7</b>	<b>8</b>	<b>15</b>
$a_2c_1$	$s_{11}$	3 ( <i>25</i> )	1 ( <i>11</i> )	2 ( <i>19</i> )
	$s_{12}$	2 ( <i>12</i> )	1 ( <i>7</i> )	3 ( <i>20</i> )
	$s_{13}$	3 ( <i>23</i> )	2 ( <i>20</i> )	1 ( <i>7</i> )
	$s_{14}$	3 ( <i>14</i> )	1 ( <i>7</i> )	2 ( <i>8</i> )
	$s_{15}$	2 ( <i>19</i> )	1 ( <i>8</i> )	3 ( <i>21</i> )
$T_i$		<b>13</b>	<b>6</b>	<b>11</b>
$a_2c_2$	$s_{16}$	1 ( <i>7</i> )	2 ( <i>8</i> )	3 ( <i>19</i> )
	$s_{17}$	3 ( <i>23</i> )	1 ( <i>7</i> )	2 ( <i>21</i> )
	$s_{18}$	2 ( <i>15</i> )	1 ( <i>3</i> )	3 ( <i>18</i> )
	$s_{19}$	1 ( <i>8</i> )	2 ( <i>14</i> )	3 ( <i>24</i> )
	$s_{20}$	2 ( <i>8</i> )	1 ( <i>7</i> )	3 ( <i>21</i> )
$T_i$		<b>9</b>	<b>7</b>	<b>14</b>

For factor  $B$ , the test statistics  $D_B^{pr.q} = 9.70$  [ $0.005 < p(\chi_{(2)}^2) < 0.01$ ] that is not significant at adjusted  $\alpha^* = \alpha/7 = 0.007$ : the application of computer science modeling does not differ with respect to the three different measuring times (for calculations see Appendix B.3).

The test statistics for factor  $B$  within factor  $C$  is  $D_{B|C}^{pr.q} = 10.40$  [ $0.034 < p(\chi_{(4)}^2) < 0.035$ ] that is not significant at adjusted  $\alpha^* = \alpha/7 = 0.007$ : the application of computer science modeling does not differ with respect to the three different measuring times determined by the two methods.

The test statistics for factor  $B$  within factor  $A$  is  $D_{B|A}^{pr,q} = 20.80$  [ $0.0001 < p(\chi_{(4)}^2) < 0.0003$ ] that is significant at adjusted  $\alpha^* = 0.007$ : the application of computer science modeling differs with respect to the three different measuring times within the two class levels.

The test statistics factor  $B$  within the factor combination  $A \times C$  is  $D_{B|A \times C}^{pr,q} = 22.80$  [ $0.003 < p(\chi_{(8)}^2) < 0.004$ ] that is significant at adjusted  $\alpha^* = 0.007$ : the application of computer science modeling differs for the two methods with respect to interaction  $A \times C$ .

The test statistics for the interaction  $B \times A$  is  $D_{B \times A}^{pr,q} = 11.10$  [ $0.0039 < p(\chi_{(2)}^2) < 0.004$ ] that is significant at adjusted  $\alpha^* = 0.007$ : the application of computer science modeling differs for the three measuring times with respect to the two grades.

The test statistics for the interaction  $B \times C$  is  $D_{B \times C}^{pr,q} = 0.70$  [ $0.70 < p(\chi_{(2)}^2) < 0.71$ ] that is not significant at adjusted  $\alpha^* = 0.007$ : the application of computer science modeling does not differ for the three measuring times with respect to the methods.

The test statistics for the second-order interaction  $A \times B \times C$  is  $D_{A \times B \times C}^{pr,q} = 1.30$  [ $0.52 < p(\chi_{(2)}^2) < 0.53$ ] that is not significant at adjusted  $\alpha^* = 0.007$ : the application of computer science modeling does not differ for the two grades with respect to measuring times and methods.

#### **Further rank order statistics for SPF- $pr \bullet q$ experimental designs**

Further *rank order statistics* for SPF- $pr \bullet q$  experimental designs could not be identified in the literature.

## **5. Conclusions**

Experiments are of high value in educational research because they are well-suited to study causal relationships. The process-product paradigm implies that researchers conduct educational experiments with respect to relevant factors that are important for teaching and learning in classroom environments: presage, context, and process variables. Data collected in educational contexts are often ordinal and do not have the properties of a normal distribution. In addition, sample sizes are often small. Thus, *rank order statistics* should be used for analysis.

Three types of factorial experimental designs have been described in this article: CRF- $pq$ , SPF- $p \bullet q$ , SPF- $pr \bullet q$ . These designs have been chosen in order (1) to analyze interactions between two or more variables and (2) to include repeated measures comparisons. For each design, appropriate *rank order statistics* have been described and exemplified.

One should be aware of the fact that this article mostly describes experimental education from a technical point of view. As a matter of course, researchers should first think about theoretical constructs and possible hypotheses before they choose experimental designs and statistical tests. Nevertheless, once an experimental

design has been reasonably chosen, experiments can be conducted with small samples and with data not meeting the requirements of parametric tests.

Subsequent endeavors should treat hierarchical experimental designs including *rank order statistics* to investigate school classes as a whole. In addition, multivariate designs and appropriate rank order statistics should be systematically applied to the field of educational research. Using the experimental designs with *rank order statistics*, researchers are enabled to conduct classroom experiments with small samples and dependent variables on an ordinal scale.

**Acknowledgements:** We thank Jasmin Pfaff for her support in promptly obtaining materials via interlibrary loan.

## 6. References:

- Akritis, M. G., Arnold, S. F., & Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, *92*, 258–265. <<http://dx.doi.org/10.1080/01621459.1997.10473623>>
- Aloe, A. M., & Becker, B.J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics*, *37*(2), 278–297. <<http://dx.doi.org/10.3102/1076998610396901>>
- Beasley, T. M. (2000). Nonparametric tests for analyzing interactions among intra-block ranks in multiple group repeated measures designs. *Journal of Educational and Behavioral Statistics*, *25*(1), 20–59.
- Bredenkamp, J. (1974). Nonparametrische Prüfung von Wechselwirkungen [In German]. *Psychologische Beiträge*, *16*, 398–416.
- Brunner, E., & Dette, H. (1992). Rank procedures for the two factor mixed model. *Journal of the American Statistical Association*, *87*, 884–888. <<http://dx.doi.org/10.1080/01621459.1992.10475292>>
- Brunner, E., Dette, H., & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, *92*, 1494–1502. <<http://dx.doi.org/10.1080/01621459.1997.10473671>>
- Brunner, E., Puri, M. L., & Sun, S. (1995). Nonparametric methods for stratified two-sample designs with application to multiclinic trials. *Journal of the American Statistical Associations*, *90*, 1004–1014. <<http://dx.doi.org/10.1080/01621459.1995.10476602>>
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Cooper, M. (1988). Nonparametric statistics. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (pp. 705–710). Oxford: Pergamon Press.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on aptitude-treatment interactions*. New York: Irvington.
- David, H. A. (1970). *Order statistics*. New York: Wiley and Sons.
- DeVaney, T. A. (2001). Statistical significance, effect size, and replication: what do the journals say? *The Journal of Experimental Education*, *69*(3), 310–320. <<http://dx.doi.org/10.1080/00220970109599490>>
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (392–431). London: Macmillan.
- Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.
- Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric statistical inference*. New York: Dekker.
- Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979–1993: The preparation of graduate students to read the educational research literature. *Educational Researcher*, *14*(2), 5–11.
- Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, *71*(1), 105–131. <<http://dx.doi.org/10.3102/00346543071001105>>

- Hedeker, D. (2008) Multilevel models for ordinal and nominal variables. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 237–275). New York: Springer.  
<[http://dx.doi.org/10.1007/978-0-387-73186-5\\_6](http://dx.doi.org/10.1007/978-0-387-73186-5_6)>
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341–370. <<http://dx.doi.org/10.3102/1076998606298043>>
- Hedges, L.V. (2011). Effect Sizes in Three-Level Cluster-Randomized Experiments. *Journal of Educational and Behavioral Statistics*, 36(3), 346–380. <<http://dx.doi.org/10.3102/1076998610376617>>
- Helmke, A. (2005). *Unterrichtsqualität. erfassen, bewerten, verbessern* [In German]. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Helmke, A. (2006). Unterrichtsforschung. In K. H. Arnold, U. Sandfuchs, & J. Wiechmann (Eds.), *Handbuch Unterricht* [In German] (pp. 56–65). Bad Heilbrunn: Julius Klinkhardt.
- Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*. Malabar, FL: Krieger.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Hox, J. J. (2000). Multilevel analysis of grouped and longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 15–32). London: Lawrence Erlbaum Associates.
- Huck, S. W. (2009). *Reading statistics*. Boston: Pearson.
- Jin, H., Barnard, J., & Rubin, D. B. (2010). A modified general location model for noncompliance with missing data: Revisiting the New York City school choice scholarship program using principal stratification. *Journal of Educational and Behavioral Statistics*, 35(2), 154–173.
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, 68, 350–386.
- Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: a methodological review. *The Journal of Experimental Education*, 69(3), 280–309.  
<<http://dx.doi.org/10.1080/00220970109599489>>
- Kirk, R. E. (1995). *Experimental design* (3<sup>rd</sup> ed.). Pacific Grove, CA: Brooks/Cole.
- Klauer, K. J. (1973). *Das Experiment in der pädagogischen Forschung* [In German]. Düsseldorf: Schwann.
- Klauer, K. J. (2005). *Das Experiment in der pädagogisch-psychologischen Forschung* [In German] (Reprint). Münster: Waxmann.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47, 583–621. <<http://dx.doi.org/10.1080/01621459.1952.10483441>>
- Kvam, P. H., & Vidakovic, B. (2007). *Nonparametric statistics with applications to science and engineering*. New York: Wiley. <<http://dx.doi.org/10.1002/9780470168707>>
- Lemmer, H. H., & Stoker, D. J. (1967). A distribution-free analysis of variance for the two-way classification. *South African Statistical Journal*, 1, 67–74.
- Lewis, D. G. (1968). *Experimental design in education*. London: University of London Press.
- Lienert, G. A. (1978). *Verteilungsfreie Methoden in der Biostatistik* [In German] (Bd. II). Meisenheim am Glan: Hain.
- Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance *F* test. *Review of Educational Research*, 66, 579–620.
- Luh, W. M. (1999). Developing trimmed mean test statistics for two-way fixed-effects ANOVA models under variance heterogeneity and nonnormality. *The Journal of Experimental Education*, 67(3), 243–264.  
<<http://dx.doi.org/10.1080/00220979909598355>>
- Luh, W. M., & Guo, J. H. (2006). Heteroscedastic test statistics for one-way analysis: the trimmed means and



- Hall's transformation conjunction. *Journal of Experimental Education*, 74(1), 75–100.  
<<http://dx.doi.org/10.3200/JEXE.74.1.75-100>>
- MacDonald, P. (1999). Power, type I, and type II error rates of parametric and nonparametric tests. *The Journal of Experimental Education*, 67(4), 367–379. <<http://dx.doi.org/10.1080/00220979909598489>>
- Marjoribanks, K. M. (1988). Interaction effects. In J. P. Keeves, *Educational research, methodology, and measurement: An international handbook* (pp. 625–628). Oxford: Pergamon Press.
- Mielke, P. W., & Berry, K. J. (1999). Multivariate test for correlated data in completely randomized designs. *Journal of Educational and Behavioral Statistics*, 24(2), 109–131.
- Moerbeek, M. (2008). Powerful and cost-efficient designs for longitudinal intervention studies with two treatment groups. *Journal of Educational and Behavioral Statistics*, 33(1), 41–61.  
<<http://dx.doi.org/10.3102/1076998607302630>>
- Mok, M. (1995). Sample size requirements for 2-level designs in educational research. *Multilevel Modelling Newsletter*, 7(2), 11–15.
- Narens, L. (2007). *Introduction to the theories of measurement and meaningfulness and the use of symmetry in science*. London: Lawrence Erlbaum Associates.
- O'Malley, P. M. (1988). Detection of interaction. In J. P. Keeves, *Educational research, methodology, and measurement: An international handbook* (pp. 625–628). Oxford: Pergamon Press.
- Pett, M. A. (1997). *Nonparametric statistics for health care research*. London: Sage.
- Pietsch, M. (2011). Fehlende Daten bei Unterrichtsbeobachtungen: Eine Sensitivitätsanalyse anhand von Daten der Schulinspektion Hamburg [In German]. *Empirische Pädagogik*, 25(1), 47–87.
- Rashid, M. M. (1996). Inference based on ranks for two-way models with a grouping factor and a repeated measure factor. *Journal of Nonparametric Statistics*, 6, 27–43.  
<http://dx.doi.org/10.1080/10485259608832661>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Raykov, T., & Marcoulides, G. A. (2010). Multivariate effect size estimation: Confidence interval construction via latent variable modeling. *Journal of Educational and Behavioral Statistics*, 35(4), 407–421.  
<http://dx.doi.org/10.3102/1076998609359792>
- Roberts, F. S. (1979). *Measurement theory with applications to decision making, utility, and the social sciences*. London: Addison Wesley.
- Rosenthal, R., & Rubin, D. B. (2003).  $r_{\text{equivalent}}$ : A simple effect size indicator. *Psychological Methods*, 8(4), 492–496. <<http://dx.doi.org/10.1037/1082-989X.8.4.492>>
- Schochet, P. Z. (2008). Statistical power for random assignment evaluations of educational programs. *Journal of Educational and Behavioral Statistics*, 33(1), 62–87. <<http://dx.doi.org/10.3102/1076998607302714>>
- Schwippert, K. (2002). Verfahren zum Ersetzen fehlender Daten – Ein anwendungsbezogenes Beispiel zum Hot-Deck-Verfahren [In German]. *Empirische Pädagogik*, 16, 509–532.
- Sen, P. K., & Puri, M. L. (1977). Asymptotically distribution-free aligned rank order tests for composite hypotheses for general multivariate linear models. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 39, 175–186.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395), 826–831. <<http://dx.doi.org/10.1080/01621459.1986.10478341>>
- Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics*, 35(1), 26–53.  
<<http://dx.doi.org/10.3102/1076998609345252>>
- Shuell, T. (1996). Teaching and learning in a classroom context. In D. C. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York: Macmillan.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics*. New York: McGraw-Hill.

- Steland, A. (1997). On a rank test in a two factor mixed model with varying dependent repeated measurements. *Journal of Nonparametric Statistics*, 8, 215–236. <<http://dx.doi.org/10.1080/10485259708832721>>
- Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *The Journal of Experimental Education*, 70(1), 80–93. <<http://dx.doi.org/10.1080/00220970109599499>>
- Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in The Journal of Experimental Education. *The Journal of Experimental Education*, 66(1), 75–83. <<http://dx.doi.org/10.1080/00220979709601396>>
- Vargha, A., & Delaney, H. D. (1998) The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, 23, 170–192.
- Wellenreuther, M. (2007). *Lehren und Lernen – aber wie?* Baltmannsweiler: Schneider.
- West, C. K., Carmody, C., & Stallings, W. M. (1983). The quality of research articles in the *Journal of Educational Researcher*, 25, 26–30.
- Wilcox, R. R. J.-H. (2006). Graphical methods for assessing effect size: some alternatives to Cohen’s d. *Journal of Experimental Education*, 74(4), 353–367.
- Wolfe, D. A., Dean, A., & Hartlaub, B. A. (1990). Nonparametric rank-based test procedures for non-additive models in the two-way layout I. No replications. *Communications in Statistics, Theory and Methods*, 19(11), 4355–4382. <<http://dx.doi.org/10.1080/03610929008830446>>
- Zendler, A., & Pfeiffer, T. (2009). Methodische Befunde zu durchgeführten Unterrichtsexperimenten [In German]. *Empirische Pädagogik*, 23(2), 208–221.
- Zhang, J. L., & Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by “death”. *Journal of Educational and Behavioral Statistics*, 28(4), 353–368. <<http://dx.doi.org/10.3102/10769986028004353>>
- Zimmermann, D. W. (1998). Increasing the power of non-parametric tests by detecting and down-weighting outliers. *The Journal of Experimental Education*, 64(1), 71–78. <<http://dx.doi.org/10.1080/00220973.1995.9943796>>

## Appendix A

*Rank order statistics* (David, 1970; Sen & Puri, 1977; Hettmansperger, 1984) are non-parametric procedures whose test statistics are calculated based on ranks derived from the original data. Well-known examples of rank order statistics are the pair-difference-test by Wilcoxon, the U-test by Mann-Whitney, the  $\chi_r^2$ -test by Friedman, the H-test by Kruskal-Wallis.

Using the simple U-test, the concept of *rank order statistics* is explained now: An educational experiment investigates the effect of two methods ( $a_1$ ,  $a_2$ ) on learning to read. Ten learners participate in the experiment. Five students learn with method  $a_1$ , five students with method  $a_2$ . Table A.1 contains the data  $y_i$ ,  $i=1, \dots, 10$ .

Table A.1 (left side) shows the original data, and table A.1 (right side) shows the ranks derived from the original data in ascending order.

**Table A.1**

*Data set for the U-test by Mann-Whitney*

data $Y$			ranks $R$	
$a_1$	$a_2$		$a_1$	$a_2$
13	1	→	8	1
9	6		6	4
16	7		9	5
20	3		10	2
11	5		7	3
<b><math>n</math></b>	<b>5</b>		<b><math>T</math></b>	<b>40</b>
			<b>15</b>	

**Table A.2**

*Abbreviated form for data sets*

$a_1$	$a_2$
8 ( <i>13</i> )	1 ( <i>1</i> )
6 ( <i>9</i> )	4 ( <i>6</i> )
9 ( <i>16</i> )	5 ( <i>7</i> )
10 ( <i>20</i> )	2 ( <i>3</i> )
7 ( <i>11</i> )	3 ( <i>5</i> )
<b><math>T</math> 40</b>	<b>15</b>

Table A.2 shows the abbreviated form used for data sets in the present article: ranks are listed in normal letters; italics in parentheses are used for original data.

The difference regarding the sums of the ranks for the two methods  $a_1$  and  $a_2$  indicates possible differences in learning to read. For method  $a_1$  one calculates a sum of ranks  $T_1 = 40$  and for method  $a_2$  one calculates  $T_2 = 15$ .

To test whether the two sum of ranks differ, the test statistics of the U-test by Mann and Whitney is used:

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_2 + 1)}{2} - T_1;$$

Where  $n_1$  and  $n_2$  are the sizes of the two individual samples, and  $T_1$  is the sum of ranks of the first sample.

With respect to the example  $U = 5 \cdot 5 + \frac{5 \cdot (5 + 1)}{2} - 40 = 0$ .

To decide, whether the U-statistics is significant, it is compared to the expected so-called *critical value*  $U_{krit(\alpha; n_1, n_2)}$  at a selected level of significance  $\alpha$ . Because  $U = 0 < U_{krit(.05; 5, 5)} = 2$ , the test statistics  $U$  is significant at  $\alpha = .05$ , or abbreviated:  $U = 0 [p(U_{(5, 5)}) < 0.05]$ .

**Appendix B**

1. The bifactorial  $H$ -test for a CRF-32 design: calculations for the example in section 4.1

**Calculating  $H_{A \times B}$**

**Table B.1**

*Data set for the bifactorial  $H$ -test ( $N=30$ )*

<i>data Y</i>						<i>ranks R</i>						
$a_1$	$a_1$	$a_2$	$a_2$	$a_3$	$a_3$	$a_1$	$a_1$	$a_2$	$a_2$	$a_3$	$a_3$	
$b_1$	$b_2$	$b_1$	$b_2$	$b_1$	$b_2$	$b_1$	$b_2$	$b_1$	$b_2$	$b_1$	$b_2$	
6	1.5	12	8	28.5	12.5	9	1	20	12	29	21	
2	6.5	11	7	29	14	2	10	18	11	30	24	
4.5	2.5	11.5	8.5	27	13.5	6	3	19	13	27	23	
3	5	10.5	9	28	13	4	7	17	14	28	22	
4	5.5	10	9.5	26.5	16	5	8	16	15	26	25	
$n_i$	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	$T_i$	<b>26</b>	<b>29</b>	<b>90</b>	<b>65</b>	<b>140</b>	<b>115</b>

$$H = \frac{12}{N \cdot (N+1)} \cdot \sum_{i=1}^{p \cdot q=3 \cdot 2} \frac{T_i^2}{n_i} - 3 \cdot (N+1) = \frac{12}{30 \cdot (30+1)} \cdot \left( \frac{26^2}{5} + \frac{29^2}{5} + \frac{90^2}{5} + \frac{65^2}{5} + \frac{140^2}{5} + \frac{115^2}{5} \right) - 3 \cdot (30+1) = 27.43$$

**Calculating  $H_A$**

**Table B.2**

*Data pooled under factor A (original data in parentheses) ( $N=15$ )*

<i>pooled data Y</i>						<i>recalculated ranks R</i>			
		$a_2$		$a_3$		$a_1$	$a_2$	$a_3$	
7.5	(6+1.5)	20	(12+8)	41	(28.5+12.5)	2	7.5	12.5	
8.5	(2+6.5)	18	(11+7)	43	(29+14)	4	6	15	
7	(4.5+2.5)	20	(11.5+8.5)	40.5	(27+13.5)	1	7.5	11	
8	(3+5)	19.5	(10.5+9)	41	(28+13)	3	9.5	12.5	
9.5	(4+5.5)	19.5	(10+9.5)	42.5	(26.5+16)	5	9.5	14	
$n_{i(a)}$	<b>5</b>	<b>5</b>		<b>5</b>		$T_{i(a)}$	<b>15</b>	<b>40</b>	<b>65</b>

$$H_A = \frac{12}{N \cdot (N+1)} \cdot \sum_{i(a)=1}^{p=3} \frac{T_{i(a)}^2}{n_{i(a)}} - 3 \cdot (N+1) = \frac{12}{15 \cdot (15+1)} \cdot \left( \frac{15^2}{5} + \frac{40^2}{5} + \frac{65^2}{5} \right) - 3 \cdot (15+1) = 12.50$$

**Calculating  $H_B$**

**Table B.3**

Data pooled under factor B (original data in parentheses) ( $N=10$ )

pooled data Y					recalculated ranks R	
		$b_2$			$b_1$	$b_2$
46.5	(6+12+28.5)	22	(1.5+8+12.5)	→	10	1
42	(2+11+29)	27.5	(6.5+7+14)		8	4
43	(4.5+11.5+27)	24.5	(2.5+8.5+13.5)		9	2
41.5	(3+10.5+28)	27	(5+9+13)		7	3
40.5	(4+10+26.5)	31	(5.5+9.5+16)		6	5
$n_{i(b)}$	<b>10</b>	<b>10</b>			$T_{i(b)}$	<b>40 15</b>

$$H_B = \frac{12}{N \cdot (N+1)} \cdot \sum_{i(b)=1}^{q=2} \frac{T_{i(b)}^2}{n_{i(b)}} - 3 \cdot (N+1) = \frac{12}{10 \cdot (10+1)} \cdot \left( \frac{40^2}{5} + \frac{15^2}{5} \right) - 3 \cdot (10+1) = 6.82$$

**Calculating  $H_{A \times B}$**

$$H_{A \times B} = H - H_A - H_B = 27.43 - 12.50 - 6.82 = 8.11$$

2. The bifactorial  $D^{p,q}$ -test for a SPF-3•2 design: calculations for the example in section 4.2

Calculating  $D_{B|A}^{p,q}$  (see Table B.4)

**Table B.4**

Data set for the bifactorial  $Dp\cdot q$ -test ( $N=15$ )

		data $Y$		ranks $R$	
		$b_1$	$b_2$	$b_1$	$b_2$
$a_1$	$s_1$	17	19	1	2
	$s_2$	14	25	1	2
	$s_3$	13	24	1	2
	$s_4$	15	14	2	1
	$s_5$	11	16	1	2
				$T_i$	<b>6</b> <b>9</b>
$a_2$	$s_6$	12	15	1	2
	$s_7$	16	13	2	1
	$s_8$	14	13	2	1
	$s_9$	15	28	1	2
	$s_{10}$	19	20	1	2
				$T_i$	<b>7</b> <b>8</b>
$a_3$	$s_{11}$	24	3	2	1
	$s_{12}$	27	6	2	1
	$s_{13}$	18	7	2	1
	$s_{14}$	29	8	2	1
	$s_{15}$	19	12	2	1
				$T_i$	<b>10</b> <b>5</b>
				$T_{i(b)}$	<b>23</b> <b>22</b>

$$D_{B|A}^{p,q} = \frac{12}{n_q \cdot q \cdot (q+1)} \sum_{i=1}^{p \cdot q = 3 \cdot 2} T_i^2 - 3 \cdot n_q \cdot p \cdot (q+1) = \frac{12}{5 \cdot 2 \cdot (2+1)} (6^2 + 9^2 + 7^2 + 8^2 + 10^2 + 5^2) - 3 \cdot 5 \cdot 3 \cdot (2+1) = 7.0$$

Calculating  $D_B^{p,q}$  (ranks added under factor A, see Table B.4)

$$D_B^{p,q} = \frac{12}{N \cdot q \cdot (q+1)} \sum_{i(b)=1}^{q=2} T_{i(b)}^2 - 3 \cdot N \cdot (q+1) = \frac{12}{15 \cdot 2 \cdot (2+1)} (23^2 + 22^2) - 3 \cdot 15 \cdot (2+1) = 0.07$$

Calculating  $D_{A \times B}^{p,q}$

$$D_{A \times B}^{p,q} = D_{B|A}^{p,q} - D_B^{p,q} = 7.0 - 0.07 = 6.93$$

3. The trifactorial  $D^{pr,q}$ -test for a SPF-22•3 design: calculations for the example in section 4.3

**Table B.5**

Data set for the trifactorial  $D^{pr,q}$ -test ( $N=20$ )

		data Y			ranks R			
		$b_1$	$b_2$	$b_3$	$b_1$	$b_2$	$b_3$	
$a_1c_1$	$s_1$	17	10	11	3	1	2	
	$s_2$	20	15	8	3	2	1	
	$s_3$	28	16	9	3	2	1	
	$s_4$	16	10	22	2	1	3	
	$s_5$	24	16	6	3	2	1	
					$T_i$	<b>14</b>	<b>8</b>	<b>8</b>
$a_1c_2$	$s_6$	13	9	23	2	1	3	
	$s_7$	4	20	24	1	2	3	
	$s_8$	4	19	25	1	2	3	
	$s_9$	14	3	19	2	1	3	
	$s_{10}$	14	17	28	1	2	3	
					$T_i$	<b>7</b>	<b>8</b>	<b>15</b>
$a_2c_1$	$s_{11}$	25	11	19	3	1	2	
	$s_{12}$	12	7	20	2	1	3	
	$s_{13}$	23	20	7	3	2	1	
	$s_{14}$	14	7	8	3	1	2	
	$s_{15}$	19	8	21	2	1	3	
					$T_i$	<b>13</b>	<b>6</b>	<b>11</b>
$a_2c_2$	$s_{16}$	7	8	19	1	2	3	
	$s_{17}$	23	7	21	3	1	2	
	$s_{18}$	15	3	18	2	1	3	
	$s_{19}$	8	14	24	1	2	3	
	$s_{20}$	8	7	21	2	1	3	
					$T_i$	<b>9</b>	<b>7</b>	<b>14</b>
					$T_{i(a)}$	<b>21</b>	<b>16</b>	<b>23</b>
					$T_{i(b)}$	<b>22</b>	<b>13</b>	<b>25</b>
					$T_{i(c)}$	<b>43</b>	<b>29</b>	<b>48</b>
					$T_{i(a)}$	<b>27</b>	<b>14</b>	<b>19</b>
					$T_{i(c)}$	<b>16</b>	<b>15</b>	<b>29</b>

**Calculating**  $D_B^{pr,q}$  (ranks added under factors A and C, see Table B.5)

$$D_B^{pr,q} = \frac{12}{N \cdot q \cdot (q+1)} \sum_{i(b)=1}^{q=3} T_{i(b)}^2 - 3 \cdot N \cdot (q+1) = \frac{12}{20 \cdot 3 \cdot (3+1)} (43^2 + 29^2 + 48^2) - 3 \cdot 20 \cdot (3+1) = 9.70$$

**Calculating**  $D_{B|C}^{pr,q}$  (ranks added under factor C, see Table B.5)

$$D_{B|C}^{pr,q} = \frac{12 \cdot r}{N \cdot q \cdot (q+1)} \sum_{i(a)=1}^{p \cdot q = 2 \cdot 3} T_{i(a)}^2 - 3 \cdot N \cdot (q+1) = \frac{12 \cdot 2}{20 \cdot 3 \cdot (3+1)} (21^2 + 16^2 + 23^2 + 22^2 + 13^2 + 25^2) - 3 \cdot 20 \cdot (3+1) = 10.40$$

**Calculating**  $D_{B|A}^{pr,q}$  (ranks added under factors A, see Table B.5)

$$D_{B|A}^{pr,q} = \frac{12 \cdot p}{N \cdot q \cdot (q+1)} \sum_{i(c)=1}^{r,q=2:3} T_{i(c)}^2 - 3 \cdot N \cdot (q+1) = \frac{12 \cdot 2}{20 \cdot 3 \cdot (3+1)} (27^2 + 14^2 + 19^2 + 16^2 + 15^2 + 29^2) - 3 \cdot 20 \cdot (3+1) = 20.80$$

**Calculating**  $D_{B|A \times C}^{pr,q}$  (see Table B.5)

$$D_{B|A \times C}^{pr,q} = \frac{12 \cdot p \cdot r}{N \cdot q \cdot (q+1)} \sum_{i=1}^{p \cdot r, q=2:2:3} T_i^2 - 3 \cdot N \cdot (q+1) =$$

$$= \frac{12 \cdot 2 \cdot 2}{20 \cdot 3 \cdot (3+1)} (14^2 + 8^2 + 8^2 + 7^2 + 8^2 + 15^2 + 13^2 + 6^2 + 11^2 + 9^2 + 7^2 + 14^2) - 3 \cdot 20 \cdot (3+1) = 22.80$$

**Derived calculations**

$$D_{B \times A}^{pr,q} = D_{B|A}^{pr,q} - D_B^{pr,q} = 20.80 - 9.70 = 11.10$$

$$D_{B \times C}^{pr,q} = D_{B|C}^{pr,q} - D_B^{pr,q} = 10.40 - 9.70 = 0.70$$

$$D_{A \times B \times C}^{pr,q} = D_{B|A \times C}^{pr,q} - D_B^{pr,q} - D_{B \times A}^{pr,q} - D_{B \times C}^{pr,q} = 22.80 - 9.70 - 11.10 - 0.70 = 1.30$$