Making statistics meaningful

The factorial survey as a teaching tool in social work education

Wallander, Lisa

2025

Document Version:
Publisher's PDF, also known as Version of record

WORKING PAPER

Nr 2 • 2025

ISSN 1650-8971

Making statistics meaningful

The factorial survey as a teaching tool in social work education

Author: **Lisa Wallander**

LUNDS UNIVERSITET

Socialhögskolan, Lunds universitet • Box 23, 221 00 Lund

Contents

Preface

Around 10 to 15 years ago, as part of a postdoctoral project at a university in southern Sweden, I explored the potential of using the factorial survey approach – a vignette experiment employing fictitious descriptions to model judgments (Rossi & Nock 1982) – as a teaching tool in higher education for social work. My primary aim was to provide students with a tool to explore and analyze the principles underlying their own individual conclusions regarding aspects such as problem identification, problem severity, and problem resolution. Given its focus on professional judgment and critical thinking, the approach could be seamlessly integrated into any teaching module addressing these competencies (see, e.g., Wallander & Molander, 2016). However, having been extensively involved in teaching statistics at both undergraduate and graduate levels, I quickly recognized that this tool also had significant potential to make statistical concepts more meaningful for students. By using students' own judgments of 100 short fictitious descriptions of individuals using alcohol or drugs as data, students would not only engage in the process of data collection but also analyze data directly related to themselves. My hope was that this approach would not only enhance student motivation but also provide insight into the social construction and social contexts of statistical data. Furthermore, the teaching tool proved to be highly adaptable, allowing for integration into the teaching of univariate, bivariate, and multivariate statistics. At that time, I had the opportunity to serve as a guest lecturer on a criminology master's course on research methods, where I was able to test these ideas with a relatively small group of students. To facilitate the exercise, I developed 100 vignettes presented through a paper-and-pencil questionnaire, an SPSS dataset containing the dimensions of the vignettes, and a manual outlining various univariate, bivariate, and multivariate statistical analyses that could be performed with the data. The exercise was well received, and the students provided positive feedback. This working paper is structured as a short "teaching note", detailing the background and the steps of the exercise along with the students' evaluations. It is intended for an academic audience who are already familiar with the statistical concepts and foundational principles of univariate, bivariate, and multivariate analysis. Having remained unpublished for too long, I hope that its inclusion in this working paper series will encourage educators — not only in social work but across the social and behavioral sciences — to integrate the factorial survey teaching tool into their statistics courses.

Lisa Wallander, April 2025 in Lund

Contact details:
E-mail: lisa.wallander@soch.lu.se

ABSTRACT

The aim of this working paper is to expand the range of pedagogical tools for teaching statistics to social work students, by introducing an innovative tool that is based on the factorial survey and intended for practicing statistical analysis. The factorial survey, which was pioneered by the American sociologist Peter Rossi around 45 years ago, involves the use of experimentally constructed fictitious descriptions (vignettes) in the study of social judgments. In the present two-day-exercise, each student first judged the severity of the alcohol or drug use described in 100 randomly sampled vignettes. By carrying out univariate, bivariate and multivariate statistical analyses, the students then identified the predictors of their own individual severity judgments. The statistics labs were instructor-led and included time both for interpreting the results from a statistical point of view and for comparing the substantive results of the students' individual analyses. A brief assessment of the exercise revealed high mean levels of perceived increases in the students' statistical knowledge. This teaching tool incorporates several key features that are considered essential for making statistics education more engaging and relevant to students. It is also proposed as an ideal method for integrating statistical instruction into substantive subject courses.

KEYWORDS: Factorial survey, vignette experiment, social work, teaching statistics, teaching tool

INTRODUCTION

Thirty-five years ago, an article entitled "The numerate social worker" was published in the Journal of Social Work Education (Taylor, 1990). The author maintained that numeracy – i.e., the capacity for quantitative thought and expression – is essential for the problem-solving activities of everyday social work practice, which include assessing clients' needs, making decisions in a context characterized by uncertainty, and monitoring change (Taylor, 1990). Since the publication of this article, the growing emphasis on using empirical evidence to inform professional practice has underscored the need for social workers to read, understand, and critically apply the findings of quantitative research, including evaluations of interventions (Teater et al., 2017). An ability to comprehend and reason using statistics also makes practitioners less susceptible to deception (Slootmaeckers et al., 2014) and less dependent on the knowledge of individuals in neighboring professions (Glisson & Fischer, 1987). As formulated by Capshew (2005, p. 857), "practicing social work without a basic understanding of statistics is like agreeing to play chess without knowing the rules: what you do not know can be used against you and your clients".

However, teaching quantitative skills and statistics to social work students has proved to be a challenge, not least due to students' lack of interest and negative feelings about the subject (e.g. Lalayants, 2012). As noted by Davis and Mirick (2015, p. 318), social work students generally believe that there is no "goodness of fit" between themselves and statistics. Because many students fail to see the relevance of statistics to the field of social work, they also tend to evaluate statistics courses as pointless (Lalayants, 2012). Furthermore, studies have shown that the well-

researched phenomenon labelled "statistics anxiety" (Baird, 2016; Chew & Dillon, 2014; Onwuiegbuzie & Wilson, 2003) is widespread among social work students, affecting as many as half (Davis & Mirick, 2015) or even two thirds (Condron et al., 2018). Social work students who experience high levels of statistics anxiety tend to avoid opportunities to use statistics (Davis & Mirick, 2015) and to fear and be less interested in general research courses (Gredig et al., 2020).

As of today, the challenges of teaching statistics to social work students are widely acknowledged, and the scientific literature includes numerous studies examining remedies and interventions designed to enhance student motivation and alleviate statistics anxiety (for overviews, see e.g. Davis & Mirick, 2017; Junius & Sidell, 2009; Lalayants, 2012). Such remedies may focus on classroom climate and instructor attributes, such as "immediacy behaviors," which encompass verbal and non-verbal cues that convey psychological and physical availability (Tonsing, 2018). Additionally, they may involve instructors' enthusiasm and passion for their subject matter, which have been shown to enhance student engagement (Smith & Martinez-Moyano, 2012; Teater et al., 2017).They also include specific teaching strategies and pedagogical tools, such as incorporating humor into the teaching of statistics (Neumann et al., 2009), using decision-making flow charts to inform the choice of inferential statistics for a particular research problem (Calderwood, 2012), incorporating online statistics labs into research methods courses (Elliott et al., 2013), and making use of real-world data from web-based databases (Bolen, 2006), community agencies (Wells, 2006) or from the students themselves (Marson, 2007). In a fairly recent effort to build upon and leverage existing knowledge on teaching statistics in higher education, Teater and colleagues (2017) developed, piloted, and evaluated a curriculum consisting of ten individual lessons on quantitative research methods for social work students in the UK.

Against this background, the aim of this working paper is to expand the repertoire of pedagogical strategies and tools in this field by introducing an innovative teaching tool. This tool is based on a vignette experiment method known as *the factorial survey* (Auspurg & Hinz, 2015; Rossi & Nock, 1982; Wallander, 2009) and is designed for practicing basic univariate, bivariate, and multivariate statistical analysis. This teaching tool incorporates several key features often cited as essential for making statistics education more engaging and meaningful for students. These include connecting to a real-world topic of interest, allowing students to collect their own data (about themselves), and offering hands-on learning opportunities (Davis & Mirick, 2017; Lalayants, 2012; Smith & Martinez-Moyano, 2012; Teater et al., 2017).

The factorial survey, introduced by American sociologist Peter Rossi in the late 1970s (Rossi, 1979; Rossi & Nock, 1982), is a hybrid method that integrates the strengths of surveys (external validity) and experiments (internal validity) to study individuals' beliefs and judgments. At the heart of the method is the *vignette* – a fictional description of a person or situation – composed of one *level* (value) from multiple *dimensions* (variables), to which individuals are asked to respond. Contemporary practice in factorial survey research usually involves analyzing the pooled judgments of a large population of individuals (Wallander, 2009). However, the method is equally suited, in principle, for modeling the judgment principles of single individuals – that is, how each person utilizes the information presented in the vignettes to form their judgments (Jasso, 1988; Wallander & Laanemets, 2017; Wallander & Molander, 2016; Wallander 2019). As such, it is highly suited to being employed in smaller groups of individuals, such as those found in

higher education. Additionally, due to the basic characteristics of the method – which involve multiple judgments of vignettes, in which a fairly large number of variables are simultaneously manipulated – it is particularly well-suited for teaching univariate, bivariate and multivariate statistical analysis.

The tool [the factorial survey teaching tool] takes the form of a two-day exercise in which each student first collects data about her own judgments – in this case about the severity of alcohol and drug use. They then conduct statistical analyses to identify the factors influencing their own judgments. Below, I will (1) introduce the factorial survey method, as employed in the teaching tool; (2) describe the statistical analyses that were conducted using the students' data in the software program SPSS (IBM Statistical Package for the Social Sciences; https://www.ibm.com/analytics/spss-statistics-software); (3) present the students' brief assessments of the exercise and; (4) discuss the benefits and limitations of the teaching tool. The data used to illustrate and assess the tool were collected in October 2013 as part of a research methods course on the master's program in criminology at a university in southern Sweden. The students taking the course came from various scientific disciplines, including social work, and may be regarded as a small convenience sample of social and behavioral science master's students at a university specializing in professional education.

CONSTRUCTING A FACTORIAL SURVEY TEACHING TOOL

In accordance with the basic principles of the factorial survey, vignettes are generated by selecting one level from a number of dimensions, which are believed to be significant determinants of the judgment of interest (Rossi & Nock, 1982; for detailed instructions, see Auspurg & Hinz, 2015). The current exercise makes use of vignettes designed to assess judgments about the severity of alcohol and drug use – a slightly modified version of those used in Samuelsson & Wallander (2014). In that study, the authors were particularly interested in examining the influence of social categories on practitioners' severity judgments. Figure 1 displays an example of a vignette:

Figure 1. An example of a vignette

> Peter is 42 years old and works as a teacher. He is a single parent. When smoking cannabis, he has several times lately ended up being confused and behaving strangely, according to his friend.

How severe do you find this person's alcohol or drug use?

Not severe at all Very severe

1 2 3 4 5 6 7

This vignette design is based on nine dimensions: (1) sex, (2) ethnicity, (3) age, (4) socio-economic status, (5) civil status, (6) presence/absence of children, (7) substance, (8) frequency of consequence, (9) type of consequence. Each of these dimensions comprises a specific number of levels. For example, "socio-economic status" contains four levels (unemployed/working class/middle class/well-established) while "substance" is represented by three levels (alcohol/cannabis/cocaine). Table 1 shows the dimensions, levels and wordings of the design, and the percentage of vignettes for each dimension level.

The complete set of all possible combinations of dimension levels represents the maximum number of unique vignettes and is sometimes referred to as the "vignette universe". In the present case the vignette universe includes 2*3*3*4*2*2*3*3*5 = 12,960 different vignettes. In a factorial survey, the respondents do not judge all of the vignettes comprising the vignette universe, but rather samples of vignettes, which are drawn from the universe by random or quota sampling (for more details on sampling in factorial surveys, see Auspurg & Hinz, 2015; Dülmer, 2016). In the present exercise, we made use of *one* random sample of 100 vignettes, which was judged by all of the students participating in the exercise. The vignettes were pre-generated by the instructor using a combination of Excel to generate random numbers and a custom-designed software program to transform these numbers into coherent texts. The Excel file should have as many rows as there are vignettes in the sample and as many columns as there are vignette dimensions, plus an additional column for the vignette ID number. For a detailed overview of random sampling in factorial survey studies, see Ludwick et al. (2004). For guidance on constructing vignettes using a combination of Excel, SPSS, and Word, please contact the author of this working paper.

Table 1. The vignette design used in the exercise (see Samuelsson & Wallander, 2014)

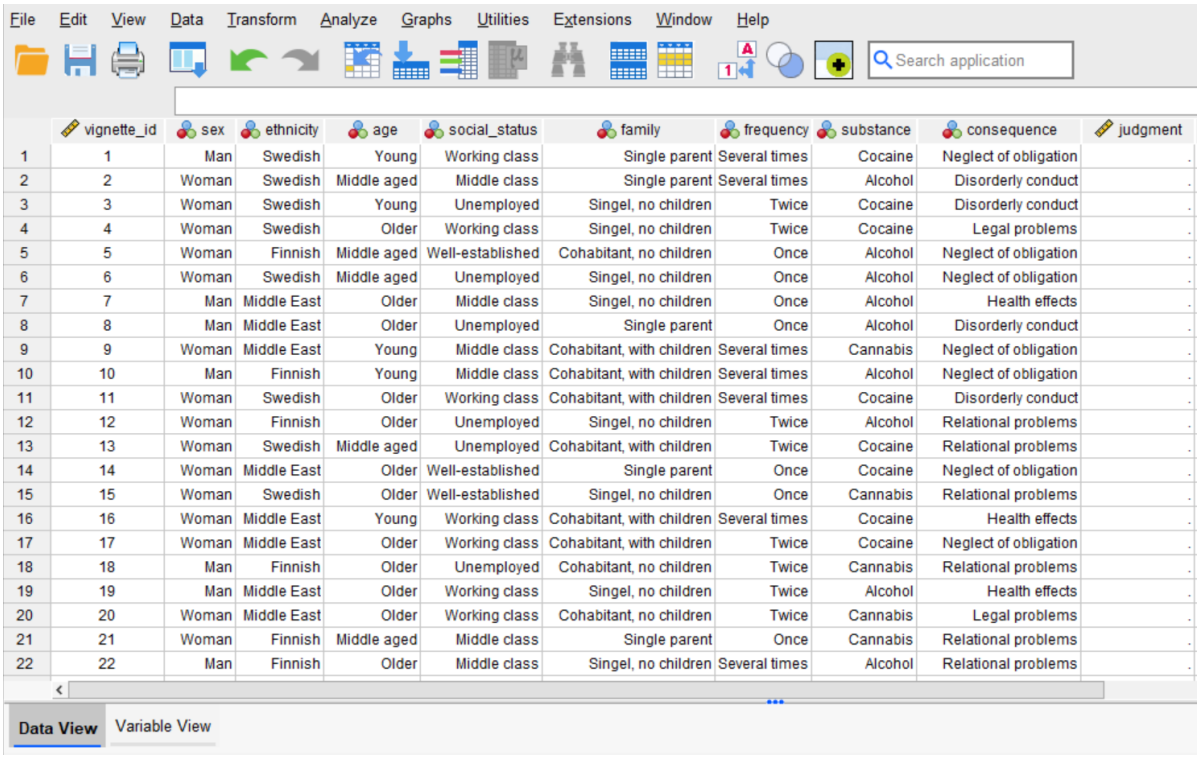
Dimension and Levels	Wordings	Vig %
Sex		
Man	Johan/Peter/Anders	18 %
Woman	Pekka/Arto/Matti	16 %
Ethnicity	Mohammed/Reza/Ramin	16 %
Swedish	Helen/Anna/Carina	15 %
Finnish	Pirjo/Sirkka/Riika	18 %
Middle East	Fatima/Soheila/Leila	17 %
Age		
Young	is 23 years old and	33 %
Middle aged	is 42 years old and	34 %
Older	is 58 years old and	33 %
Socio-economic status		
Unemployed	unemployed.	25 %
Working class	works as a cleaner/mental hospital nurse.	25 %
Middle class	works as a teacher/journalist.[1]	25 %
Well-established	works as a lawyer/advanced engineer.[1]	25 %
Civil status		
Single		50 %
Cohabitant	He/she is single/a single parent.	50 %
Children	He/she lives with his/her partner/and children.	
No		50 %
Yes		50 %
Substance		
Alcohol	When drinking alcohol, he/she has	33 %
Cannabis	When smoking cannabis, he/she has	34 %
Cocaine	When taking cocaine, he/she has	33 %
Frequency of consequence		
Once	recently, on one occasion	33 %
Twice	a couple of times during the last year	34 %
Several times	several times lately	33 %
Type of consequence		
Relational problems	ended up in a quarrel with a friend.	20 %
Neglect of obligation	missed an important meeting the following day.	20 %
Health effects (alc.)	ended up not remembering what happened last night.	20 %
Health effects (can.)	ended up being confused and behaving strangely, according to his/her friend.	
Health effects (coc.)	ended up feeling such anxiety that he/she wasn't able to handle his/her ordinary commitments.	
Disorderly conduct	been barred from a restaurant because of disorderly conduct.	20 %
Legal problems	been arrested by the police.	20 %

[1]When the person in the vignette was young, he/she was described as a student of these occupations.

EMPLOYING THE TOOL FOR TEACHING STATISTICS

At the start of the exercise, each student gathers data about their own judgments. In this case, each student assessed the severity of alcohol or drug use described in 100 vignettes (see Figure 1 above for an example of such a vignette). This was done as homework, prior to getting together for the statistics labs. The time required to make these judgments usually varies between 40 and 90 minutes (Wallander & Molander, 2016). The statistics labs, conducted over two days, were instructor-led, with the instructor using a data file containing her own judgments as an example. Students also had access to a brief instruction manual outlining the SPSS commands used in the statistical analyses. Alternatively, any general statistics book covering SPSS commands, such as Pallant (2020), could be used as a reference. At the start of the first statistics lab, the students were given a prepared SPSS data file. This file was identical to the Excel file generated during the random number creation process, containing the values—along with added labels—that represent the dimension levels defining the 100 vignettes. The students constructed an additional variable representing their vignette judgments and entered their judgments of the 100 vignettes into the designated rows in the data view (see Figure 2).

Figure 2. An illustration of the prepared SPSS data file



	vignette_id	sex	ethnicity	age	social_status	family	frequency	substance	consequence	judgment
1	1	Man	Swedish	Young	Working class	Single parent	Several times	Cocaine	Neglect of obligation	.
2	2	Woman	Swedish	Middle aged	Middle class	Single parent	Several times	Alcohol	Disorderly conduct	.
3	3	Woman	Swedish	Young	Unemployed	Singel, no children	Twice	Cocaine	Disorderly conduct	.
4	4	Woman	Swedish	Older	Working class	Singel, no children	Twice	Cocaine	Legal problems	.
5	5	Woman	Finnish	Middle aged	Well-established	Cohabitant, no children	Once	Alcohol	Neglect of obligation	.
6	6	Woman	Swedish	Middle aged	Unemployed	Singel, no children	Once	Alcohol	Neglect of obligation	.
7	7	Man	Middle East	Older	Middle class	Singel, no children	Once	Alcohol	Health effects	.
8	8	Man	Middle East	Older	Unemployed	Single parent	Once	Alcohol	Disorderly conduct	.
9	9	Woman	Middle East	Young	Middle class	Cohabitant, with children	Several times	Cannabis	Neglect of obligation	.
10	10	Man	Finnish	Young	Middle class	Cohabitant, with children	Several times	Alcohol	Neglect of obligation	.
11	11	Woman	Swedish	Older	Working class	Cohabitant, with children	Several times	Cocaine	Disorderly conduct	.
12	12	Woman	Finnish	Older	Unemployed	Singel, no children	Twice	Alcohol	Relational problems	.
13	13	Woman	Swedish	Middle aged	Unemployed	Cohabitant, with children	Twice	Cocaine	Relational problems	.
14	14	Woman	Middle East	Older	Well-established	Single parent	Once	Cocaine	Neglect of obligation	.
15	15	Woman	Swedish	Older	Well-established	Singel, no children	Once	Cannabis	Relational problems	.
16	16	Woman	Middle East	Young	Working class	Cohabitant, with children	Several times	Cocaine	Health effects	.
17	17	Woman	Middle East	Older	Working class	Cohabitant, with children	Twice	Cocaine	Neglect of obligation	.
18	18	Man	Finnish	Older	Unemployed	Cohabitant, no children	Twice	Cannabis	Relational problems	.
19	19	Man	Middle East	Older	Working class	Singel, no children	Twice	Alcohol	Health effects	.
20	20	Woman	Middle East	Older	Working class	Cohabitant, no children	Twice	Cannabis	Legal problems	.
21	21	Woman	Finnish	Middle aged	Middle class	Single parent	Once	Cannabis	Relational problems	.
22	22	Man	Finnish	Older	Middle class	Singel, no children	Several times	Alcohol	Relational problems	.

Data View Variable View

Entering the judgments into the program generally takes less than 30 minutes. The data were then ready to be analyzed using univariate, bivariate and multivariate statistics (see Table 2).

Table 2. The main forms of analysis covered in the exercise

	Main forms of analysis
Univariate statistics	Frequency tables Measures of central tendency Measures of variability
Bivariate statistics	Independent-Samples T-test One-Way ANOVA Eta Square (Measure of effect size) Bivariate linear regression analysis
Multivariate statistics	Multiple linear regression analysis (MRA) – main effects Multiple linear regression analysis (MRA) – interaction effects Calculating predicted values

Univariate statistics

The choice of univariate statistics will be based on the characteristics of the variables, i.e., their level of measurement. In this exercise, students' vignette judgments (rated on a scale from 1 to 7; ordinal data treated as an equidistant continuous variable) were analyzed using measures of central tendency and variability. Meanwhile, the vignette dimensions (nominal or ordinal data) were examined using frequency tables. The frequency tables indicated that the vignette dimensions were well-balanced, meaning that the dimension levels were equally represented in the vignette sample. This balance was achieved through the randomization process (see Table 1 above for the distribution of vignette dimensions). The results depicting the distribution of the students' judgments highlighted their individual tendencies in assessing the vignettes as more or less severe. Naturally, this tendency varied among participants in the exercise. Our example student in this working paper (see more below) had an average severity rating of 5.13 (SD = 1.00), whereas another student's average rating was as low as 3.97 (SD = 1.42).

Bivariate statistics

In the same way, the choice of bivariate analysis will be dependent on the characteristics of the variables. Since this exercise involves a continuous dependent variable (the student's severity judgments) and several categorical independent variables (the vignette dimensions), we compared the arithmetic means of each student's judgments across the vignette dimensions. The Independent-Samples T-test was used for dichotomous vignette dimensions, while One-way ANOVA was applied to vignette dimensions with three or more levels. In addition, we used Excel to compute Eta Square – a measure of effect size – for each vignette dimension (for formulas, see Pallant, 2020, pp. 255, 268). As a final stage of bivariate analysis, we computed a series of nine bivariate linear regressions, where each independent variable was represented by one or a set of dummy variables (i.e. binary – 0 or 1 – variables). This was a very useful way of demonstrating that different methods of bivariate analysis yield approximately the same results.

For examples, see Table 3, which displays two bivariate relationships between vignette dimensions (children and type of consequence) and one student's judgments about severity.

Table 3. A selection of results from bivariate analyses predicting the effects of the vignette dimensions on one student's severity judgments (n=100)

	Independent-Samples T-test	One-Way ANOVA (incl. Tukey test)	Linear regression analysis
Bivariate analysis			
Children (No = ref)	Mean diff: -0.500		b = 0.500
Yes	t = -2.565		t = 2.565
	p = .012		p = .012
	Eta^2 = 0.063		R^2 = 0.063
Type of consequence (Relational problems = ref)		Mean diff (p):	b (p):
Neglect of obligation		-0.650 (.091)	0.650 (.013)
Health effects		-1.750 (.000)	1.750 (.000)
Disorderly conduct		-0.500 (.297)	0.500 (.054)
Legal problems		-1.250 (.000)	1.250 (.000)
		F = 14.139	F = 14.139
		Eta^2 = 0.373	R^2 = 0.373

The Eta^2 and R^2 values, which are identical, reflect the influence of each vignette dimension on the judgments. These values can be used to rank the dimensions according to their overall significance as predictors of a student's severity judgments. For our example student, the ranking of vignette dimensions in accordance with their influence on severity judgments, from the strongest to the weakest impact, was as follows: type of consequence (R^2 = 0.373), frequency of consequence (R^2 = 0.186), substance (R^2 = 0.125), presence or absence of children (R^2 = 0.063), ethnicity (R^2 = 0.047), socio-economic status (R^2 = 0.030), age (R^2 = 0.004), sex (R^2 = 0.001), and civil status (R^2 = 0.000).

The mean differences from the T-test and One-Way ANOVA, along with the unstandardized regression coefficients from the bivariate linear regression analyses, provide insights into how the student assigned weight to the respective dimension level when making severity judgments of the 100 vignettes. In the given example, the student perceived the situation as more severe when the individual in the vignette was a parent rather than not (b = 0.500). Regarding the type of consequence, the student rated health effects as the most severe (b = 1.750), followed by legal problems (b = 1.250), neglect of obligation (b = 0.650), disorderly conduct (b = 0.500), with relational problems serving as the reference category (equals b=0.000). Since the data pertains to a single student and does not represent a sample from a larger population, significance testing with p-values cannot be used to test null hypotheses about a broader population. Instead, they serve as indicators of the consistency and robustness of the judgment principles applied by this student.

Multivariate statistics

In this exercise, multivariate regression analysis was employed to examine the main and interactive effects of the vignette dimensions on the students' judgments. Table 4 displays the results from a multivariate linear regression analysis, in which one student's judgments have been regressed on nine vignette dimensions.

Table 4. A selection of results from multivariate analyses predicting the effects of the vignette dimensions on one student's severity judgments (n=100)

Multivariate analysis	b unstandardized coefficient	p-value
Constant	3.240	.000
Sex		
(Man = ref)		
Woman	-0.035	.794
Ethnicity		
(Swedish = ref)		
Finnish	0.062	.718
Middle East	0.154	.370
Age		
(Young = ref)		
Middle aged	0.062	.730
Older	0.037	.835
Socio-economic status		
(Unemployed = ref)		
Working class	-0.068	.730
Middle class	0.159	.418
Well-established	0.037	.851
Civil status		
(Single = ref)		
Cohabitant	-0.081	.544
Children		
(No = ref)		
Yes	0.283	.045
Substance		
(Alcohol = ref)		
Cannabis	0.301	.081
Cocaine	0.955	.000
Frequency of consequence		
(Once = ref)		
Twice	0.639	.000
Several times	0.733	.000
Type of consequence		
(Relational problems = ref)		
Neglect of obligation	0.579	.014
Health effects	1.556	.000
Disorderly conduct	0.489	.041
Legal problems	1.340	.000
R^2	0.667	

All vignette dimensions are dummy coded, meaning that a binary (0 or 1) variable has been created for each level of the dimension, except for one, which serves as the reference category. Since each dimension of the vignettes was generated using random numbers, the bivariate correlations between the independent variables are very low (not shown here), ensuring that the data maintain a high level of experimental control. Consequently, the relationships between the vignette dimensions and severity judgments will change only minimally when statistical control is applied in a multivariate linear regression analysis. Indeed, a comparison of the results from the bivariate and multivariate analyses for the dimensions presence of children and type of consequence (see Tables 3 and 4) reveals only minor numerical differences and no changes in the substantive findings. One important value of interest is R^2, which represents the percentage of variance in severity judgments explained by the vignette dimensions. A high R^2 value indicates that the student consistently took these dimensions into account when making severity judgments, whereas a low R^2 value suggests that the dimensions had a less systematic influence. For instance, our example student had an R^2 value of 0.667, while another student's value was as low as 0.296.

After analyzing the main effects of the vignette dimensions on severity judgments, the students manually calculated their individual predicted values for a few "typical" individuals using alcohol or drugs. These calculations allowed them to compare their predicted severity judgments for the selected typical cases. Let us take the following vignette example: "Riika is 42 years old and works as a journalist. She lives with her partner and children. When smoking cannabis, she has a couple of times during the last year missed an important meeting the following day". For our example student, the predicted severity rating in this case would be calculated as follows: 3.240 + (-0.035) + 0.062 + 0.062 + 0.159 + (-0.081) + 0.283 + 0.301 + 0.639 + 0.579 = 5.209 (Note: The rating scale ranged from 1 to 7). As the final step in the multivariate analysis, we created a set of interaction terms using dummy coding (e.g., the interaction between sex and ethnicity involved five dummy variables and one reference category) and incorporated them as independent variables in a new multivariate linear regression analysis (not presented here).

Each stage of analysis was accompanied by a discussion in which we interpreted the results using statistical terminology, and where the students were also given the opportunity to compare and discuss their results, which displayed different aspects of their individual judgment principles based on this particular vignette design. For more details about how to make sense of such results in the context of professional judgments, see Wallander and Molander (2016), Wallander and Laanemets (2017) and Wallander (2019).

AN ASSESSMENT OF THE FACTORIAL SURVEY TEACHING TOOL

At the end of the second day of instruction, the students were asked to assess the different parts of the exercise, in terms of (1) their perceived level of knowledge prior to the exercise (on a scale from 1 = no knowledge to 5 = considerable knowledge), and (2) their perceived increase in knowledge as a result of the exercise (on a scale from 1 = no increase to 5 = substantial increase). As the two-day exercise was part of a compulsory research methods course, while participation in this research project was voluntary, the students were given the opportunity to abstain from

assessing the exercise, and one student out of 12 chose to do so. The remaining 11 students received information about the study and about their individual rights as participants in a scientific study. Having received this information, they signed and handed in a written statement of informed consent. It is also worth noting that students submitted their assessments online and anonymously. Additionally, the author/instructor served as a guest lecturer for two days in the course and did not act as an examiner for this or any other course in the master's program in criminology. The results of the students' assessments are shown in Table 5.

Table 5. Student assessment of the exercise, in terms of the perceived level of knowledge prior to the exercise[1] and the perceived increase in knowledge as a result of the exercise[2] (n=11)

Form of Statistical Analysis	Prior knowledge		Knowledge increase	
	Mean	Min-Max	Mean	Min-Max
Univariate analysis				
Frequency tables	2.91	1-5	3.45	1-5
Measures of central tendency	2.91	1-5	3.27	1-5
Measures of variability	2.64	1-5	3.82	2-5
Bivariate analysis				
Independent-Samples T test	2.27	1-4	3.91	2-5
One-Way ANOVA	1.55	1-4	4.27	2-5
Eta Square (Measure of effect size)[3]	1.30	1-3	4.45	3-5
Bivariate regression analysis	2.00	1-4	4.09	2-5
Multivariate analysis				
MRA, linear (main effects)	2.18	1-4	4.09	2-5
MRA, linear (interaction effects)	1.73	1-3	3.91	2-5
Calculating predicted values	1.73	1-3	4.18	2-5

[1] Perceived level of knowledge prior to the exercise (1 = no knowledge – 5 = considerable knowledge)
[2] Perceived increase in knowledge as a result of the exercise (1 = no increase – 5 = substantial increase)
[3] Due to item nonresponse, n=10 on the "prior knowledge" item

The results presented in Table 5 are fairly uniform in the sense that the mean levels of perceived previous knowledge are located at the lower end of the scale (between 1 and 3), while the mean levels of perceived knowledge gain can be found at the upper end of the scale (between 3 and 5). In addition, while the students in general had higher levels of previous knowledge about univariate analysis than about bivariate and multivariate analysis, the perceived increases in the levels of knowledge were generally higher for bivariate and multivariate analysis than for univariate analysis.

DISCUSSION

The aim of the current working paper has been to present an innovative and creative teaching tool – the factorial survey teaching tool – which can be effectively utilized for teaching basic univariate, bivariate and multivariate statistics. The exercise presented above was carried out in a group of 12 students on the master's program in criminology at a university in southern Sweden. Since these students came from diverse academic disciplines and universities, their levels of

statistical knowledge varied considerably. While some had taken research courses that covered all of the material included in the exercise, others had no previous experience of working with statistical software. Taking this variation into account, the exercise worked well both as a means of refreshing existing knowledge of basic univariate, bivariate and multivariate statistics and as an introduction to these methods of analysis. This suggests that the tool could also be used in teaching at the undergraduate level. Moreover, although this particular exercise constituted part of a research methods course, the concise format of the factorial survey teaching tool makes it ideal for the teaching of statistics in substantive social work courses. In fact, it has been suggested that the incorporation of statistics modules into substantive subject courses could positively benefit students' development and their retention of statistical literacy skills (Davis & Mirick, 2017; Hageman & Pecukonis, 2020; Slootmaeckers et al., 2014; Taylor, 1990). Such a strategy would naturally require the creation of factorial survey vignettes specifically designed to assess the predictors of a judgment relevant to the course in question. Fortunately, the factorial survey has been increasingly employed for the study of professional judgments in social work (Taylor, 2006; Wallander, 2012), and there are quite a few vignette designs available (e.g., Mc Elhinney et al., 2021; Mullineux et al., 2020; Wallander & Blomqvist, 2019; Wallander & Laanemets, 2017), which can easily be adapted and employed as teaching tools in social work education.

One limitation of this teaching tool, stemming from the fact that the majority of the data consists of randomly generated numbers, is that it cannot be used to explore the complexities that typically characterize real-world data. For instance, while the data are suitable for practicing regression analysis with sets of dummy variables, they cannot be used for regression diagnostics or for identifying mediator variables. However, if the vignette design includes two different judgments per vignette, one of these judgments could potentially be treated as a mediating variable influencing the effects of the vignette dimensions on the other judgment (cf. Wallander & Molander, 2014). On the plus side, the teaching tool offers valuable opportunities to illustrate the principles of experimental design. For example, by analyzing the distributions of the vignette dimensions (which will be approximately or entirely balanced) and the bivariate correlations between these dimensions (which will be relatively low and, in a larger random sample of vignettes, nearly orthogonal), students can examine the connections between randomization, experimental control, and causality. Furthermore, beyond helping students become familiar with the basics of the factorial survey method, this teaching tool can be easily expanded to include a written component focused on argument analysis, using Stephen Toulmin's (1958) widely recognized model of argumentation (Wallander & Molander, 2016).

As noted above, the factorial survey teaching tool was favorably assessed by the students in terms of the knowledge gained by participating in the exercise. However, these results should be interpreted with care, since the students' estimations are measures of subjective knowledge. In addition, in the absence of a pretest and control group, and considering the fact that the sample is very small, the results of the assessment should be treated as exploratory findings. Future studies should include objective measures of knowledge gain, as well as a pretest and posttest with a control group to evaluate these findings. That said, the high average levels of perceived increases in the students' statistical knowledge could be seen as an indication of a highly positive experience with practicing statistical analysis. Such experiences could help to alter students'

potentially negative attitudes and beliefs about numeracy (Bolin et al., 2012; Lalayants, 2012; Taylor, 1990) and to alleviate problems with statistics anxiety (Condron et al., 2018).

As the title of this working paper indicates, the present author suggests that the factorial survey teaching tool constitutes a way of *making statistics meaningful* to students of social work. One reason for this is that it uses the students themselves as the source from which the data are collected. Over and above this relatively common strategy for increasing the likelihood that students will actually care about the statistics being examined (e.g. Marson, 2007), the factorial survey teaching tool involves another, even more inspiring, source of "meaningfulness". Since the students are likely unable to fully track the manipulation of dimensions within the vignettes – given the large number of vignettes and dimensions – the analytical exercises may reveal judgment patterns that students themselves are unaware of (Wallander, 2012). To take one example, the results from this particular exercise may reveal that your judgments about the severity of alcohol or drug use are dependent on the sex, ethnicity, age or socio-economic status of the user – which you may not have been consciously aware of. Such results are thought-provoking for the individual student, and they also provide fertile ground for interesting discussions on core topics of social work.

References

Auspurg, Katrin & Hinz, Thomas (2015) *Factorial survey experiments* (Series: Quantitative applications in the social sciences). Sage Publications.

Baird, Stephanie L (2016) Conceptualizing anxiety among social work students: Implications for social work education. *Social Work Education*, *35*(6), 719-732. https://doi.org/10.1080/02615479.2016.1184639

Bolen, Rebecca M (2006) Utilizing web-based databases to introduce social work content in research statistics courses. *Social Work Education*, *25*(1), 17-27. https://doi.org/10.1080/02615470500477805

Bolin, Brien L; Lee, Kyoung Hag; GlenMaye, Linnea F & Yoon, Dong Pil (2012) Impact of research orientation on attitudes toward research of social work students. *Journal of Social Work Education*, *48*(2), 223-243. https://doi.org/10.5175/JSWE.2012.200900120

Calderwood, Kimberly A (2012) Teaching inferential statistics to social work students: A decision-making flow chart. *Journal of Teaching in Social Work*, *32*(2), 133-147. https://doi.org/10.1080/08841233.2012.670065

Capshew, Thomas F (2005) Motivating social work students in statistics courses. *Social Work Education*, *24*(8), 857-868. https://doi.org/10.1080/02615470500342207

Chew, Peter K & Dillon, Denise Bridget (2014) Statistics anxiety update: Refining the construct and recommendations for a new research agenda. *Perspectives on Psychological Science*, *9*(2), 196-208. https://doi.org/10.1177%2F1745691613518077

Condron, Dennis J; Becker, Jacob H & Bzhetaj, Linda (2018) Sources of students' anxiety in a multidisciplinary social statistics course. *Teaching sociology*, *46*(4), 346-355. https://doi.org/10.1177%2F0092055X18780501

Davis, Ashley & Mirick, Rebecca G (2015) MSW students' perceptions of relevance and application of statistics: Implications for field education. *Journal of Teaching in Social Work*, *35*(3), 317-336. https://doi.org/10.1080/08841233.2015.1040531

Davis, Ashley & Mirick, Rebecca G (2017) Teaching statistics to MSW students: Comparing credit and non-credit options. *Advances in Social Work*, *18*(2), 437-455. https://doi.org/10.18060/21448

Dülmer, Hermann (2016) The factorial survey: Design selection and its impact on reliability and internal validity. *Sociological Methods & Research*, *45*(2), 304-347. https://doi.org/10.1177%2F0049124115582269

Elliott, William; Choi, Eunhee & Friedline, Terri (2013) Online statistics labs in MSW research methods courses: Reducing reluctance toward statistics. *Journal of Social Work Education*, *49*(1), 81-95. https://doi.org/10.1080/10437797.2013.755095

Glisson, Charles & Fischer, Joel (1987) Statistical training for social workers. *Journal of Social Work Education*, *23*(3), 50-58. https://doi.org/10.1080/10437797.1987.10672084

Gredig, Daniel; Heinsch, Milena & Bartelsen-Raemy, Annabelle (2020) Exploring social work students' attitudes toward research courses: Comparing students in Australia and Switzerland. *Social Work Education*, *41*(4), 451–471. https://doi.org/10.1080/02615479.2020.1849086

Hageman, Sally Anne & Pecukonis, Edward V (2020) Public health financial social work: A model syllabus for research practice for graduate social work students. *Social Work Education*, *41*(3), 272-285. https://doi.org/10.1080/02615479.2020.1826920

Jasso, Guillermina (1988) Whom shall we welcome? Elite judgments of the criteria for the selection of immigrants. *American Sociological Review*, *53*(6), 919-32. https://doi.org/10.2307/2095900

Junius, Premalatha & Sidell, Nancy (2009) Dispelling fear and loathing: Engaging mathematically challenged students to learn statistics. *Journal of Baccalaureate Social Work*, *14*(2), 49-61. https://doi.org/10.18084/basw.14.2.3m3756374813t082

Lalayants, Marina (2012) Overcoming graduate students' negative perceptions of statistics. *Journal of Teaching in Social Work*, *32*(4), 356-375. https://doi.org/10.1080/08841233.2012.705259

Ludwick, Ruth; Wright, Marion Elizabeth; Zeller, Richard Allen; Dowding, Dawn W; Lauder, William & Winchell, Janice (2004) An improved methodology for advancing nursing research: Factorial surveys. *Advances in Nursing Science*, *27*(3), 224-238. no DOI

Marson, Stephen M (2007) Three empirical strategies for teaching statistics. *Journal of Teaching in Social Work*, *27*(3-4), 199-213. https://doi.org/10.1300/J067v27n03_13

Mc Elhinney, Helena; Taylor, Brian J & Sinclair, Marlene (2021) Judgements of health and social care professionals on a child protection referral of an unborn baby: Factorial survey. *Child Abuse & Neglect*, *114*, 104978. https://doi.org/10.1016/j.chiabu.2021.104978

Mullineux, Judith; Taylor, Brian J & Giles, Melanie (2020) Professional judgement about re-offending: Factorial survey. *Journal of Social Work*, *20*(6), 797-816. https://doi.org/10.1177%2F1468017319848889

Neumann, David L; Hood, Michelle & Neumann, Michelle M (2009) Statistics? You must be joking: The application and evaluation of humor when teaching statistics. *Journal of Statistics Education*, *17*(2). https://doi.org/10.1080/10691898.2009.11889525

Onwuegbuzie, Anthony J & Wilson, Vicki A (2003) Statistics anxiety: Nature, etiology, antecedents, effects, and treatments: A comprehensive review of the literature. *Teaching in higher education*, *8*(2), 195-209. https://doi.org/10.1080/1356251032000052447

Pallant, Julie (2020) *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th edition). Open University Press, McGraw-Hill Education.

Rossi, Peter H (1979) Vignette analysis: Uncovering the normative structure of complex judgments. In Robert K Merton, James S Coleman, & Peter H Rossi (Eds.), *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* (pp. 176-186). Free Press.

Rossi, Peter H & Nock, Steven L (1982) *Measuring social judgments: The factorial survey approach*. Sage Publications.

Samuelsson, Eva & Wallander, Lisa (2014) Disentangling practitioners' perceptions of substance use severity: A factorial survey. *Addiction Research & Theory*, *22*(4), 348-360. https://doi.org/10.3109/16066359.2013.856887

Slootmaeckers, Koen; Kerremans, Bart & Adriaensen, Johan (2014) Too afraid to learn: Attitudes towards statistics as a barrier to learning statistics and to acquiring quantitative skills. *Politics*, *34*(2), 191-200. https://doi.org/10.1111/1467-9256.12042

Smith, Amy E & Martinez-Moyano, Ignacio J (2012) Techniques in teaching statistics: Linking research production and research use. *Journal of Public Affairs Education*, *18*(1), 107-136. https://doi.org/10.1080/15236803.2012.12001674

Taylor, Brian J (2006) Factorial surveys: Using vignettes to study professional judgement. *British Journal of Social Work*, *36*(7), 1187-1207. https://doi.org/10.1093/bjsw/bch345

Taylor, Francis A (1990) The numerate social worker. *Journal of Social Work Education*, *26*(1), 25-35. https://doi.org/10.1080/10437797.1990.10672131

Teater, Barbra; Roy, Jessica; Carpenter, John; Forrester, Donald; Devaney, John & Scourfield, Jonathan (2017) Making social work count: A curriculum innovation to teach quantitative research methods and statistical analysis to undergraduate social work students in the United Kingdom. *Journal of Teaching in Social Work*, *37*(5), 422-437. https://doi.org/10.1080/08841233.2017.1381216

Tonsing, Kareen N (2018) Instructor immediacy and statistics anxiety in social work undergraduate students. *Social Work Education*, *37*(2), 223-233. https://doi.org/10.1080/10691898.2010.11889495

Toulmin, Stephen (1958/2003) *The uses of argument*. Cambridge University Press.

Wallander, Lisa (2009) 25 years of factorial surveys in sociology: A review. *Social Science Research*, *38*(3), 505-520. https://doi.org/10.1016/j.ssresearch.2009.03.004

Wallander, Lisa (2012) Measuring social workers' judgements: Why and how to use the factorial survey approach in the study of professional judgements. *Journal of Social Work*, *12*(4), 364-384. https://doi.org/10.1177%2F1468017310387463

Wallander, Lisa (2019) Fiktiva fall - fast inte som förr. Faktoriella surveymetoden i högre utbildning. *Pedagogisk Forskning i Sverige*, *24*(1), 46-67.

Wallander, Lisa & Blomqvist, Jan (2019) Fat words or useful concepts? Consensus and variation in conceptions of problematic drinking in contemporary Sweden. *Contemporary Drug Problems*, *46*(3), 241-264. https://doi.org/10.1177%2F0091450919852774

Wallander, Lisa & Laanemets, Leili (2017) Factorial survey in mixed methods research: Modelling professionals' individual judgements. In *SAGE Research Methods Cases*. https://www.doi.org/10.4135/9781473950146

Wallander, Lisa & Molander, Anders (2014) Disentangling professional discretion: A conceptual and methodological approach. *Professions & Professionalism*, *4*(3), 1-19. https://doi.org/10.7577/pp.808

Wallander, Lisa & Molander, Anders (2016) Learning to reason: The factorial survey as a teaching tool in social work education. *Nordic Social Work Research*, *6*(3), 234-250. https://doi.org/10.1080/2156857X.2016.1192560

Wells, Melissa (2006) Teaching notes: Making statistics "real" for social work students. *Journal of Social Work Education*, *42*(2), 397-404. https://doi.org/10.5175/JSWE.2006.200400466