



Article

Scand J Work Environ Health 2006;32(6):463-472

doi:10.5271/sjweh.1051

Games researchers play—extreme-groups analysis and mediation analysis in longitudinal occupational health research

by [Taris TW](#), [Kompier MAJ](#)

Affiliation: Radboud University Nijmegen, Department of Work and Organizational Psychology, PO Box 9104, NL-6500 HE Nijmegen, Netherlands. T.Taris@psych.ru.nl

Refers to the following texts of the Journal: [2003;29\(1\):1-4](#)
[2004;30\(4\):293-298](#) [2004;30\(6\):486-496](#) [2004;30\(6\):450-458](#)

The following articles refer to this text: [2006;32\(6\):413-419](#);
[2011;37\(4\):259-261](#); [2015;41\(1\):43-53](#)

Key terms: [extreme-groups analysis](#); [longitudinal study](#); [mediation analysis](#); [occupational health research](#); [researcher](#); [statistical power](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/17173202



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

Games researchers play—extreme-groups analysis and mediation analysis in longitudinal occupational health research

by Toon W Taris, PhD,¹ Michiel AJ Kompier, PhD¹

Taris TW, Kompier MAJ. Games researchers play—extreme-groups analysis and mediation analysis in longitudinal occupational health research. *Scand J Work Environ Health* 2006;32(6, special issue):463–472.

Objectives The study of causal processes using a longitudinal design is often hampered by two methodological problems. First, the lagged effects of a predictor variable on an outcome variable tend to be weak after control for a previous measure of this outcome. One approach that is advocated when effects are weak is to increase the extremeness of the study groups; this step often increases the significance and sizes of effects. Second, causal links are often mediated through third variables, and thus relatively complex mediational analyses are needed to understand the causal processes underlying particular associations. The present paper shows whether and when these two approaches are useful in longitudinal research.

Methods The two approaches were evaluated using data from a three-wave study among 1251 newcomers from various Western countries (mean age 20.6 years, 59% female).

Results Although the significances and effect sizes indeed increased with increasing extremeness of the study groups, extreme-groups analysis in the context of a longitudinal design may grossly bias findings. Cross-sectional applications of mediation analysis cannot provide evidence for any mediational model. Longitudinal models are better suited for examining mediation.

Conclusions Rather than using extreme-groups analysis to obtain significant effects across time, researchers should maximize the amount of change in their data by focusing on groups for which change can be expected. Especially multiphase longitudinal data sets offer good opportunities for analyzing mediation models.

Key terms extreme-groups analysis; longitudinal study; mediation analysis; statistical power.

Many occupational health researchers are strongly interested in the explanation of across-time change in the phenomena of central interest in our discipline (1). The strongest design for the study of the causal processes responsible for across-time change involves the random assignment of participants to experimental and control conditions and controlled manipulation of the variable of interest. However, in a field context, such a strong design is often not feasible. In such cases, longitudinal (prospective) research designs (involving at least two measurements of the same concept for the same subjects) are indispensable for examining causal processes. One major strength of this design is that it allows us to test whether a particular variable predicts the scores for an outcome variable as measured at a later point in time, while controlling for the base level scores for this outcome (typically measured on the same occasion as the predictor variable). As inclusion of the time-1 measurement of the outcome variable partials out the across-time stability in its time-2 measure, this design allows the

predictors of the remaining outcome change to be studied. Furthermore, as the temporal order of the predictor and the outcome is clear, this design provides a strong base for causal inferences concerning the process of interest (2).

Utilization of a longitudinal design does all but guarantee that our understanding of causal processes is enhanced. Our paper deals with two, often ill-understood, methodological problems that hamper the study of such processes, even when a longitudinal design is used. The first is that the concepts of interest may largely be stable across time, meaning that our predictors often fail to account for a significant amount of variance in the outcome variable. This is partly a matter of statistical power [ie, the probability that a statistical test rejects the null hypothesis given that the alternative hypothesis is true (3)], and the first part of this paper discusses one way in which researchers may attempt to increase the power of their tests (ie, enhancing exposure contrast by applying an extreme-groups approach).

¹ Radboud University Nijmegen, Nijmegen, Netherlands.

Reprint requests to: Professor TW Taris, Radboud University Nijmegen, Department of Work and Organizational Psychology, PO Box 9104, NL-6500 HE Nijmegen, Netherlands. [E-mail T.Taris@psych.ru.nl]

The second problem is that the causal process linking outcome A to predictor B is often complicated in that all sorts of variables may mediate the effect of B on A. For example, high job demands may lead causally to stress, in turn, causally leading to ill health. Then stress acts as a mediator of the relationship between demands and ill health. Therefore, if we are to understand precisely how job demands are linked to worker health, we must take stress into account as well. One well-used analytical procedure for studying mediation processes has been proposed by Baron & Kenny (4). The second part of this paper discusses their procedure, showing that the evidence resulting from it is inconclusive unless a longitudinal design is used.

In addressing these two issues, our contention is that the strategies to resolve these two problems are often very useful. However, we also believe that, in other cases, they fail to enhance the understanding of causal processes and that, in these cases, applications of these procedures are basically just games that researchers play to suggest profundity in their analyses. In this sense, the goal of this contribution is to stimulate occupational health researchers to think critically about the statistical recipes they apply and the way they can (or cannot) boost the findings of their longitudinal studies.

Study 1: the extreme-groups game

In occupational health research, it is often convenient to explore the relation between variables A and B using a two-stage design (5, 6). In the first stage, measures of the first variable (A) are obtained for a large sample of persons from a population of interest. In the second stage, the participants are selected on the basis of extreme scores for A (most commonly, upper and lower tertiles or quartiles). The relationship between A and B is then examined for these extreme scoring persons (eg, using logistic regression analysis or an analysis of variance). For example, a study on the association between exposure to urban nitrogen dioxide pollution and the risk of myocardial infarction may contrast patients being treated for first-time myocardial infarction to controls reporting no chest pain complaints during a first interview; those without myocardial infarction who reported any chest pain complaints are thus excluded from the study (7). Similarly, one may examine the effects of environmental annoyance on performance only for people scoring at the extreme ends of an environmental annoyance scale (8). We refer to these and similar sampling procedures as extreme-groups analyses (6).

Extreme-groups analysis was originally developed to reduce the sample size necessary to observe an ef-

fect without compromising statistical power (5, 6). Given a fixed sample size, the analysis improves cost-efficiency by allowing researchers to selectively sample the regions of the A distribution that will maximize the power of subsequent statistical tests (and the chances of observing a particular effect of interest), by maximizing the contrast between the groups to be compared. This power argument has, by and large, become the primary reason for the use of extreme-groups analysis (6), as this design allows researchers to examine even relatively weak effects without having to collect large quantities of data.

Extreme-groups analysis in longitudinal designs

One typical finding in longitudinal research is that the focal concepts are often remarkably stable across time, due to the relatively short time intervals that are commonly applied in these designs. As part of the across-time change in the outcome variable will be random error due to unreliability or unmeasured factors, our predictors will not always account for a significant proportion of the remaining change. Essentially, this is a matter of statistical power; if it is assumed that our predictors are indeed linked to the outcome variable in the population, our statistical tests are too weak to detect small across-time effects of the predictor variables on the outcome.

One way of countering this problem is to increase sample size, but this approach is often not feasible. A more practical possibility would seem to be the extreme-groups approach. This approach can be an efficient way to maximize the amount of exposure contrast in a prospective study. For example, a first round of large-scale data collection can serve the purpose of identifying low-versus high-risk groups for a particular outcome; these prototypical groups are then followed across time. In this way, one can increase statistical power within a fixed budget. In practice, researchers may be tempted to apply a variation on the extreme-groups approach. If sampling from the extreme ends of a distribution yields a power benefit and makes it easier to find significant effects, why not apply this procedure after all of the data have been collected? For instance, in a study on the relationship between job demands and ill health, one might discard the middle of the distribution of demands and compare only the participants with extreme scores on demands. Obviously, in this case, the cost-efficiency argument for conducting an extreme-groups analysis no longer applies; instead the presumed gain in power is, in itself, often tempting enough to use this approach [also called posthoc subgrouping (6)]. As across-time effects are usually weak, posthoc subgrouping would seem even more attractive in longitudinal research.

Yet there is reason to doubt whether posthoc subgrouping is effective in reaching its goal of increasing statistical power and would thus give more insight into the process under study. When one discards part of one's data, sample size decreases. If it is assumed that the scores in between the extreme categories also bear information on the association between the criterion variable and the variable used for posthoc subgrouping, the result is information loss and, hence, loss of statistical power. A second problem (that applies to extreme-groups analysis as well) is that selecting extreme groups may increase the chances that regression to the mean will occur and thus bias the results of subsequent longitudinal analyses either in favor or against the hypothesis to be tested (6, 9). Extreme-groups analysis and posthoc subgrouping both rest on the assumption that extreme scores in the sample represent the extremes of the true score distribution in the population. However, the cases in the extremes may be there due to incidental factors (eg, measurement unreliability). As these factors may be absent at a later point in time, cases with extreme scores at one point in time may, at a later point in time, well obtain scores that are closer to the sample mean.

Extreme-groups analysis of the determinants of learning among newcomers

The discussed procedures can be applied and evaluated using data from a three-wave study on the effects of organizational socialization practices on worker learning among 1215 newcomers. Although an extensive discussion of the theoretical background of this issue goes beyond our aims, a brief outline of the ideas underlying our application would seem justified. Basically, we assumed that learning—defined as the acquisition of new skills and knowledge—is partly determined by situational characteristics [including job characteristics such as demands and control (10)], as well as organizational efforts to stimulate worker learning behavior. As regards the latter, it may be expected that new workers in organizations that offer their novices many opportunities for learning will, in time, indeed acquire more skills than

others. Psychological contract theory suggests that this relationship may be mediated through the match between expected and actual opportunities for learning. That is, newcomers who are pleasantly surprised by the fact that the organization they work for provides them with better learning opportunities than they initially expected will be more motivated to benefit from these opportunities than workers who feel that their initial expectations were not met (11).

Method

Participants. The data were collected among 2509 newcomers from seven countries (12). Participants were employed for 3 to 9 months at the beginning of the study. The data collection for the second and third wave of the study occurred 1 and 2 years after the first round of data collection, respectively. The response rate in the third wave of the study, relative to the first wave, was 59%. After listwise deletion of missing values, the final sample included 1251 participants [mean age 20.6 (SD 3.5) years, 59% female].

Measures. This data set was used to examine how the quality of the organization's newcomer socialization program affected the degree to which the participants were surprised by their opportunities to extend their skills and their learning behavior. The "quality of the organization's newcomer socialization program" (organizational quality) was measured by a 4-item scale, including "I have the opportunity to move from one job to another to learn new skills". "Surprise regarding one's learning opportunities" was measured using a single item, "In my present job the opportunity to learn new things is . . ." (1 = much worse than expected, 5 = much better than expected). Learning behavior was measured using a 3-item scale, including "I have developed more knowledge and skill in tasks critical to my work unit's performance". All of the items were answered on 5-point scales. Table 1 presents the intercorrelations, coefficient alphas, means, and standard deviations for the study variables.

Table 1. Correlations, means, standard deviations and reliabilities (Cronbach's alpha, in italics) for the study variables (N=1251, all correlations significant at $P < 0.001$).

	<i>M</i>	<i>SD</i>	(1)	(2)	(3)	(4)	(5)	(6)
(1) Time-1 organizational quality	3.14	0.94	<i>0.62</i>					
(2) Time-1 surprise	3.59	1.00	0.39	<i>a</i>				
(3) Time-1 learning	3.54	0.82	0.40	0.30	<i>0.61</i>			
(4) Time-2 surprise	3.55	0.99	0.21	0.28	0.18	<i>a</i>		
(5) Time-2 learning	3.58	0.81	0.22	0.15	0.35	0.31	<i>0.66</i>	
(6) Time-3 learning	3.61	0.81	0.21	0.15	0.33	0.22	0.43	<i>0.65</i>

^a Measured with a single item: alpha could not be computed.

Statistical analysis

Table 2 presents the results of a 2 (organizational quality: low versus high) by 2 (time: time-1 versus time-2 learning behavior) analysis of variance, with repeated measures for time. For the present study, we compared the findings for four different selections of newcomers. The first selection included all of the participants (N=1251); time-1 organizational quality was dichotomized according to its median value in order to have about 50% of the participants in both groups. The second selection included only participants belonging to the upper and lower tertiles of time-1 organizational quality; in other words, the participants with the 33% intermediate scores were omitted from the sample. The third selection consisted of all newcomers obtaining scores in the upper and lower quartiles of organizational quality; thus half of the initial sample (ie, those with intermediate scores) was discarded. The fourth selection included all of the participants with scores in the 10% upper or lower ends of time-1 organizational quality. Thus table 2 allows us to see how the findings varied for different posthoc subgroupings. In addition to the means and standard deviations, table 2 presents the *F*-values and the effect sizes (partial eta-squared, reflecting the proportion of the variability in learning that is accounted for by the respective effect) for the main effects of time, organizational quality, and the time × organizational quality interaction.

Results

If being in a high-quality organization (ie, with a good newcomer socialization program) is indeed conducive to learning, newcomers in a high-quality organization should report increasing levels of learning across time. First consider the *F*-values at the bottom of table 2. Whereas the main effect of organizational quality is

highly significant, there is only a weak main effect for time; the time × organizational quality interaction effect is also significant. This pattern of effects largely applies to all four analyses, although the main effect of time ceases to be significant for the more extreme groupings. However, the effects are stronger (as evidenced by higher *F*-values and eta-squares) for the 33%, 25%, and 10% subgroups than for the full sample. Consistent with the rationale behind extreme-groups analysis (increase in statistical power), the effect sizes are largest for the most extreme grouping (ie, the 10% posthoc subgrouping), although the *F*-values are somewhat lower for this posthoc subgrouping than for the 33% and 25% posthoc subgroupings—here the smaller *N* seems to be taking its toll.

Inspection of the means also reveals an interesting pattern. Contrary to the hypothesis to be tested, the significant time × organizational quality interaction is *not* due to a stronger increase in learning for the high organizational quality group than for the low quality group. Rather, the difference between the low and high quality groups tends to decrease across time for all four analyses, whereas this tendency becomes stronger with an increasing extremeness of the posthoc subgrouping. For instance, for the full sample, the difference between the low and high organizational quality groups is 0.50 at time 1 and 0.27 at time 2 for the full sample, the corresponding figures are 1.27 and 0.61 for the 10% posthoc subgroups.

Discussion: can less be more?

Extreme-groups analysis and its posthoc counterpart seem indeed to be effective in raising the power of statistical tests. However, our example suggests that major problems (especially bias in the form of regression to the mean) may occur when this approach is used to

Table 2. Effects of varying degrees of group extremeness on the means and effect sizes for the effects of organizational quality on newcomer learning. (PSG = posthoc grouping)

Organizational quality	50% (N=1251)		33% PSG (N=936)		25% PSG (N=631)		10% PSG (N=224)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time 1								
Overall	3.55	0.83	3.55	0.86	3.56	0.90	3.47	1.05
High quality	3.80	0.74	3.88	0.74	3.98	0.70	4.14	0.64
Low quality	3.30	0.84	3.24	0.86	3.13	0.89	2.87	0.99
Time 2								
Overall	3.61	0.81	3.61	0.84	3.56	0.86	3.49	0.93
High quality	3.75	0.75	3.78	0.77	3.77	0.79	3.81	0.80
Low quality	3.48	0.85	3.44	0.87	3.37	0.89	3.20	0.95
<i>F</i> (quality) (η^2)	111.99 ^a (.08)		130.16 ^a (.12)		146.51 ^a (.19)		113.28 ^a (.34)	
<i>F</i> (time) (η^2)	6.39 ^b (.01)		2.64 (.00)		.11 (.00)		.00 (.00)	
<i>F</i> (quality × time) (η^2)	18.93 ^b (.02)		23.50 ^b (.03)		32.71 ^b (.05)		19.47 ^b (.08)	
<i>df</i> (<i>F</i>)	(1,1249)		(1,934)		(1,629)		(1,222)	

^b = *P*<0.05.
^a = *P*<0.01.

analyze longitudinal data. Our findings confirm earlier warnings that extreme-groups analysis and posthoc subgrouping may yield substantial bias (6, 9). It should be noted that the degree to which regression to the mean occurs will depend somewhat on the degree to which measurement error is responsible for the extreme scores of the cases in the extreme groups. If the concepts of interest have been measured reliably, regression to the mean should be less of a problem than if measurement unreliability is high. Thus the degree to which regression to the mean occurs (and, hence, the degree to which extreme-groups analysis and posthoc subgrouping yields misleading results) will depend partly on the nature and measurement of the concepts under study.

Furthermore, whereas extreme-groups analysis and posthoc subgrouping may be instrumental in obtaining significant P-values, the primary focus of research should be to determine what the data tell us about the phenomenon of interest (ie, effect size and practical relevance). Although extreme-groups analysis and posthoc subgrouping may be useful in exploratory and pilot research (ie, in answering the question of whether there are grounds for believing that two variables are associated), subsequent research should focus on obtaining unbiased estimates of the associations between variables in the population. In this sense, extreme-groups analysis and posthoc subgrouping are often inappropriate. Note that the degree to which extreme-groups analysis and posthoc subgrouping yield misleading results depends strongly on the amount of exposure contrast between the extreme groups. The more extreme the cases being sampled, the larger the deviation between the sample and the population estimates. For example, in the present application, we sampled 224 extreme cases from a "population" of 1251 newcomers, the result being substantial bias. However, this bias would have been much larger had we sampled 224 extreme cases from a population of 20 000 newcomers—in the latter case, the amount of contrast in our sample would have been much greater than in the first case and would have led to a higher statistical power and more bias in our estimates. In summary, discarding already collected data for the sake of obtaining significant effects may well yield biased results and, therefore, severely limit one's opportunities to generalize the findings to the population. Thus we must conclude that having fewer data does definitely not result in more insight; the extreme-groups game is often not worth playing.

Study 2: the shell game

Causal processes are often complicated in that the causal effect of one variable on another variable may be pre-

sumed to be mediated by other variables. An indepth examination of this causal process thus requires that the precise links among these variables be tested to see whether the presumed causal chain holds up. Baron & Kenny (4) proposed a simple procedure to test whether a particular variable (B) (eg, stress) acts as a mediator of the effect of variable A (demands) on C (health complaints). If B indeed mediates the effect of A on C, the following assumptions need to be satisfied: (i) the association ac between variables A and C is statistically significant; (ii) A and B are related; (iii) B is significantly related to C, after control for A; and (iv) the association ac between A and C is weaker when B is controlled, compared with the situation when B is not controlled. If ac ceases to be significant after control for B, B *fully mediates* the relationship between A and C; if ac is weaker but still significant, B *partially mediates* the relationship between A and C (4). The effects in the mediational models are usually estimated using some form of regression analysis (13).

Estimating a mediational model would seem to require a longitudinal design per definition. After all, in any sequence $A > B > C$, mediator B must truly be an independent variable relative to outcome variable C, the implication being that B must precede C in time; similarly, predictor A must truly be an independent variable relative to mediator B (14). Thus time must elapse for one variable to have an effect on another, and, therefore, a longitudinal design is indispensable to the study of mediational processes. Consider the three-phase model presented in figure 1. Although A, B, and C may well mutually influence each other, this design allows us to test the three central links of the mediation model, namely, whether there are causal links between (i) A and C (by regressing C3 on A1, controlling C1), (ii) A and B (by regressing B2 on A1, controlling B1), and B and C (by regressing C3 on B2, controlling C1). If mediation applies, the effect of A1 on C3 (controlling C1) should disappear (full mediation) or become weaker after controlling B2 (14).

In practice multiphase longitudinal studies are relatively rare in occupational health research. However, even a relatively modest two-wave study may help us

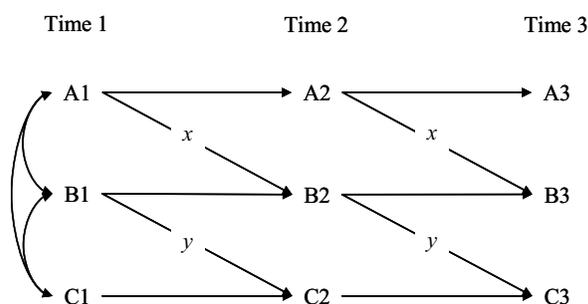


Figure 1. Three-phase mediation model.

examine mediational processes. Again, consider the model presented in figure 1. As mediation is a causal chain involving at least two causal relations (ie, $A > B$ and $B > C$), these causal relations can be tested separately using only two phases of data. For example, the causal effects of demands on learning can be tested by examining the effects of time-1 demands on time-2 learning, controlling for time-1 learning. Similarly, the effects of learning on efficacy can be tested by regressing time-2 efficacy on time-1 learning, controlling for time-1 efficacy. Partial mediation applies if both links of the presumed causal chain $A > B > C$ are confirmed; the product of the two respective lagged effects (ie, x and y in figure 1) provides an estimate of the strength of the mediational effect (14). Full mediation cannot be examined in a two-phase design; with only two phases of data, it is obviously impossible to test whether the relation between A_1 and C_3 is fully mediated by B_2 .

Interestingly, many applications of the Baron & Kenny approach do not draw on longitudinal data, but rely on single-phase (cross-sectional) data instead. This strategy is fraught with difficulties. Chief among these is that the three variables in a cross-sectional mediation study can be arranged in ($3 \times 2 \times 1 =$) six possible sequences. As the data were collected at one point in time, our design cannot help us in sorting out which causal sequences are plausible and which are not (2). In this light, it is interesting to note that researchers typically test only one causal sequence that fits the theory being tested; the five other sequences are neglected. However, this reasoning is often rickety and unjustified. First, we often conduct research because we want to explore unknown territory. Thus, it is difficult to argue a priori that one particular causal order applies while others are irrelevant; if the correct causal order were already known, our research would be superfluous. Second, in occupational health research, it is often not only conceivable, but also plausible that concepts mutually influence each other; in other words, the relationships among variables cannot adequately be conceptualized as one-way streets (15, 16). For example, high levels of social support may well lead to lower levels of stress; but it seems equally likely that stressed workers will put off their colleagues and, therefore, lead to lower levels of social support. In summary, it is often difficult to argue that, theoretically, one and only one causal order applies; in other words, the results of cross-sectional mediation analyses should be approached with caution.

Mediation analysis of the relationship between organizational quality and learning

In this section, we examine whether the relationship between organizational quality (ie, the quality of the organization's newcomer socialization program) and

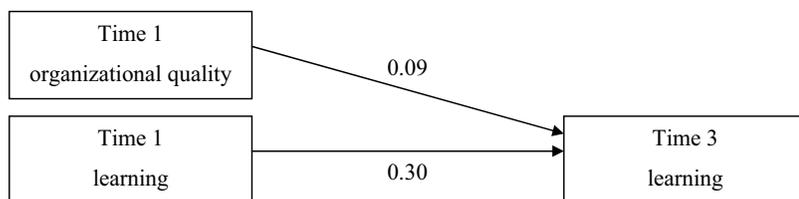
learning behavior is mediated through the degree to which newcomers are pleasantly surprised by the opportunities for learning in their organization. We present three sets of analyses. The first draws on the full three-wave study, the second uses data from the first two study waves only, whereas the third set of analyses employs only data from the first wave. In this vein, it is possible to examine whether, and to what degree, the availability of more groups of data enhances the strength of the conclusions that can be drawn regarding the mediational process that underlies the data.

Mediation in a three-wave study

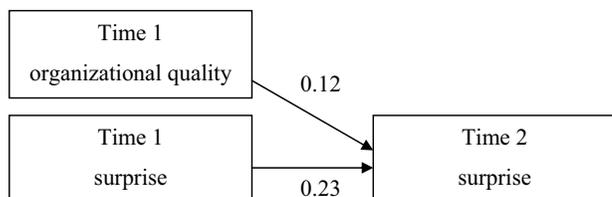
Using data from our three-wave data set on the antecedents of learning among newcomers, we examined whether time-1 organizational quality affects time-3 learning behavior and whether this relationship is mediated through time-2 surprise. Figure 2 presents the results of the respective tests.

Section A of figure 2 shows that time-1 organizational quality longitudinally predicts time-3 learning, after control for time-1 learning. Although this effect is not strong (a standardized effect of 0.09), it is in the expected direction in that newcomers who felt that their organization attached much importance to learning new skills indeed reported relatively often that they had acquired such skills. To examine whether this association was mediated through the degree to which the newcomers were pleasantly surprised by their opportunities to learn new skills, we must show that (i) time-1 organizational quality longitudinally predicts time-2 surprise (section B of figure 2), that (ii) time-2 surprise affects time-3 learning (section C of figure 2), and that (iii) the association between time-1 organizational quality and time-3 learning becomes weaker after control for time-2 surprise (section C). Figure 2 shows that these conditions are all met. Those who reported that their organization had an explicit socialization program for newcomers were indeed more pleasantly surprised by their learning opportunities. Furthermore, section C of figure 2 shows that the newcomers who were pleasantly surprised by their learning opportunities tended to report that they had learned more, controlling for time-1 learning. Although there is still a significant positive association between time-1 organizational quality and time-3 learning of 0.07, this association is lower than the association found in section A of this figure (of 0.09). The mediation effect is equal to the product of the associations between time-1 organizational quality and time-2 surprise (of 0.12) and time-2 surprise and time-3 learning (of 0.16); thus the mediation effect equals 0.02. As the association between time-1 organizational quality and time-3 learning was 0.09, about $(0.02/0.09 \times 100 =)$ 22% of the total association between

A: The association between organizational quality and learning



B: The association between organizational quality and surprise



C: The association between organizational quality and learning mediated through surprise

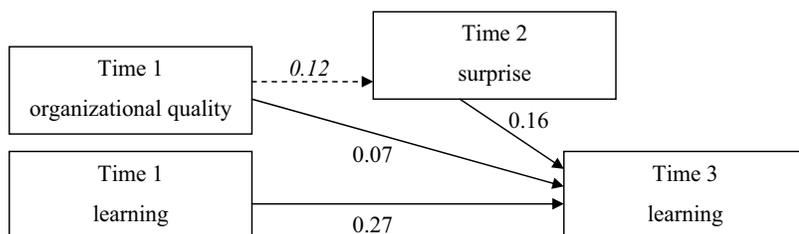


Figure 2. Standardized effects for the relations between organizational quality, surprise, and learning among newcomers: mediation examined in a three-phase model. (all estimates significant at $P < 0.01$)

time-1 organizational quality and time-3 learning is mediated through surprise (14).

Mediation in a two-wave study

To what degree can these ideas be tested in a two-wave data set? To examine this issue, we restricted our mediation analyses to the first two phases of the newcomer data set only. Two separate regressions were conducted, one with time-2 learning as the outcome and time-1 learning and time-1 surprise as predictors, and the other with time-2 surprise as the criterion variable and time-1 surprise and time-1 organizational quality as predictors. Figure 3 presents the standardized regression effects, showing that high organizational quality leads longitudinally to higher levels of surprise, whereas higher levels of surprise lead to higher levels of learning (a very weak effect of 0.05). Multiplication of the standardized effects of organizational quality on surprise (of 0.12) with the effect of surprise on learning (of 0.05) gives an estimate of the degree to which surprise mediates the relationship between organizational quality and learning,

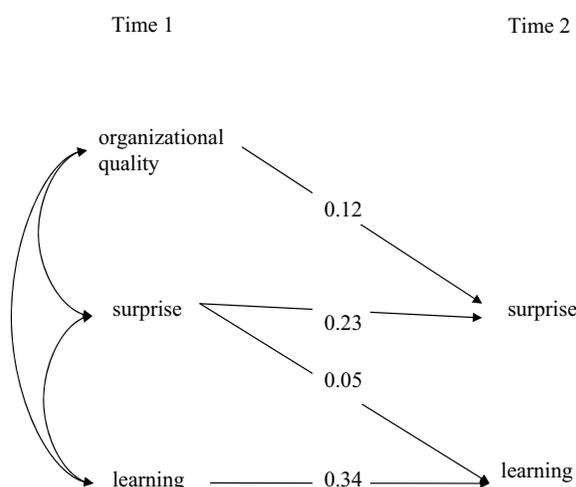


Figure 3. Standardized effects for the relations between organizational quality, surprise, and learning among newcomers: mediation examined in a two-phase model. (all estimates significant at $P < 0.05$).

yielding a not particularly impressive effect of 0.006. This figure can be compared with the estimate for this effect obtained in the three-phase analysis already

presented (of 0.02) (14), the findings suggesting that the two-phase analysis strongly underestimated an already weak mediation effect. Thus, although a two-phase study may give some indication of the presence of mediation effects, findings obtained in a three-wave study are more trustworthy.

Mediation in a cross-sectional study

To take our point that more data phases increase the trustworthiness of findings further, we tested several competing mediation models using only one set of data. This procedure corresponds with the common situation in which researchers examine mediational processes in a cross-sectional study. The mediation model that corresponds the most closely with the notions outlined is the model in which the association between organizational quality and learning is mediated through surprise (model A in figure 4). However, other causal sequences could apply as well. For example, newcomers in a high-quality organization may learn more than others, on the basis of which they may express higher levels of surprise (model B in figure 4). Thus here the relationship between organizational quality and surprise is mediated through learning. Finally, one could argue that high levels of learning lead newcomers to experience high levels of surprise; on the basis of their pleasant learning experiences in this organization, they may evaluate the quality of the organization more positively (model C in figure 4). Here the relationship between learning and organizational quality is mediated through surprise (ie, model C is the precise opposite of model A). (For simplicity, we restricted ourselves to a comparison of these three sequences only).

Figure 1 reveals that our analyses provide support for all three mediational processes discussed. (Note that our tests of the three other sequences supported these as well; the results can be obtained from the first author). Clearly, these simple analyses cannot unambiguously disentangle the relationships among the concepts

in question, at least not in the absence of more information about the correct causal order (4, 14). This situation underlines Lykken’s position (17) that “showing that one’s theory is compatible with the trends of one’s data is only weak corroboration for the theory. Showing that our theory fits the data better than all plausible alternative models, on the other hand, is strong corroboration . . . [p 34].” Given the temporal indeterminacy of cross-sectional data sets, it is unlikely that cross-sectional analyses will help researchers show that one particular mediation model fits the data better than competing models.

Discussion

Mediational processes can be examined using the procedures proposed by Baron & Kenny (4), but, in the absence of longitudinal data, the concepts of interest can be shuffled around in a random fashion, similar to what happens in the shell games played by swindlers attempting to separate unsuspecting tourists from their money. The results of the cross-sectional version of the mediation game are about as trustworthy as those of the infamous shell game—you may think you know where the ball is, but chances are that you are dead wrong. It takes a stainless-steel don’t worry, be happy attitude to put any faith in cross-sectional mediational analyses. The situation is different when one has access to a longitudinal data set. Two-wave studies offer some indication of the presence and direction of a potential mediational process. The full Baron & Kenny model can be tested when three or more waves are available, yielding the best estimate of the strength of a mediational process.

Concluding remarks

A longitudinal design is a researcher’s strongest tool for examining causality in a nonexperimental context. However, even such designs often do not yield unequivocal

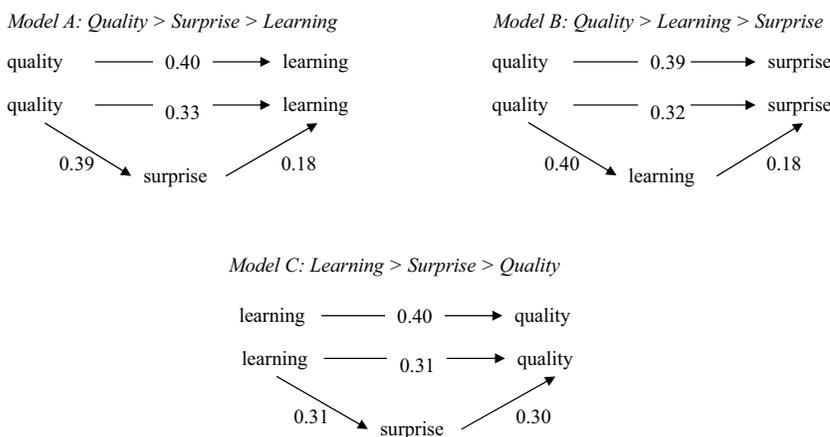


Figure 4. Comparison of the results of three sets of cross-sectional mediational models for the relationships between organizational quality, surprise, and learning among newcomers. (all effects are standardized and significant at P<0.001).

evidence regarding the causal processes of interest. Lagged effects are often weak or insignificant, while causal relationships are often presumed to be mediated through additional variables. We addressed two approaches to these problems, the extreme-groups approach (6) and the mediation analyses proposed by Baron & Kenny (4). These strategies are often applied in cross-sectional research, and it would seem tempting to probe their applicability in longitudinal designs as well.

Extreme-groups analysis is often applied to maximize the power of statistical tests by increasing the exposure contrast between the groups to be compared (6). Although this procedure would seem attractive in a longitudinal context, with its often weak effects, possible drawbacks of this approach include (i) lower sample size, resulting in a loss of statistical power, and (ii) possible bias due to regression-to-the-mean effects (14). Our empirical application suggested that the gain in statistical power by contrasting extreme groups may prevail over the loss in power due to lower sample size. Although one might argue that our findings are specific to the data set under study, theoretically the finding that extreme-groups analysis in longitudinal research may magnify bias due to regression to the mean and similar artifacts should generalize to other data sets as well (2, 6, 9). All in all, whereas extreme-groups analysis may yield an acceptable first answer to the question of whether two phenomena are related, we believe that its use in a longitudinal context should strongly be discouraged.

Instead, we advise researchers to think critically about their study design, preferably in an early stage of their research. The use of extreme-groups analysis is often instigated by the wish to detect weak longitudinal effects. These effects are usually weak due to the absence of meaningful change during the interval between the phases of the study; many participants (especially the older and more-experienced among them) will have reached a steady state as regards their work behavior, feelings, attitudes and the like, and, for these workers, little across-time change can be expected. Obviously, fancy statistical tricks cannot help much in the absence of change. Rather, we propose that researchers should maximize the chances of observing significant substantive change during their study. Potentially interesting groups include workers in the process of reorganization, newcomers, and workers that have otherwise experienced substantive change as regards their work characteristics or work situation in general (1). It seems likely that workers in the process of change will experience across-time change regarding other variables as well, and we believe these naturally occurring experiments hold considerable promise for occupational health research.

As regards the Baron & Kenny procedure to examine mediation processes (4), use of their strategy in cross-sectional studies is usually inconclusive in that it is empirically not possible to distinguish well between the different causal orders of the variables of interest (14). In this respect, mediation analysis of cross-sectional data is a suspicious shell game, with researchers trying to lure their audience into the belief that their results unveil a significant part of reality—this, in spite of the fact that their data would be consistent with many competing causal orders. Causality may well be in the eye of the beholder, but it would be wise to refrain from drawing causal inferences if the processes underlying particular patterns of associations cannot be tested adequately. In contrast, longitudinal designs offer considerably better opportunities for testing mediational processes. As is often the case, more ambitious (ie, multi-phase) longitudinal designs are more useful than simpler (two-phase) designs, in that multi-phase designs allow researchers to test several extra assumptions that underlie the use of this approach. [See Cole & Maxwell (14) for a discussion.] Yet even a limited longitudinal design presents a major step forward as compared with cross-sectional design in examining mediational processes.

References

1. Taris TW, Kompier MAJ. Challenges in longitudinal designs in occupational health psychology [editorial]. *Scand J Work Environ Health*. 2003;29(1):1–4.
2. Taris TW. A primer in longitudinal data analysis. London: Sage; 2000.
3. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum; 1988.
4. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51:1173–82.
5. Alf EF, Abrahams NM. The use of extreme groups in assessing relationships. *Psychometrika*. 1975;40:563–72.
6. Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychol Methods*. 2005;10:178–92.
7. Grazuleviciene R, Maroziene L, Dulskiene V, Malinauskiene V, Azaraviciene A, Laurinaviciene D, et al. Exposure to urban nitrogen dioxide pollution and the risk of myocardial infarction. *Scand J Work Environ Health*. 2004;30(4):293–98.
8. Österberg K, Persson R, Karlson B, Ørbæk P. Annoyance and performance of three environmentally intolerant groups during experimental challenge with chemical odors. *Scand J Work Environ Health*. 2004;30(6):486–96.
9. Campbell DT, Kenny DA. *A primer on regression artifacts*. New York (NY): Guilford Press; 1999.
10. Karasek R, Theorell T. *Healthy work: stress, productivity, and the reconstruction of working life*. New York (NY): Basic

- Books; 1990.
11. Rousseau DM. Schema, promise and mutuality: the building blocks of the psychological contract. *J Occup Organ Psychol*. 2001;74:511–41.
 12. Taris TW, Feij JA. Learning and strain among newcomers: a three-wave study on the effects of job demands and job control. *J Psychol*. 2005;138:543–63.
 13. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variables effects. *Psychol Methods*. 2002;7:83–104.
 14. Cole DA, Maxwell SE. Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *J Abnorm Psychol*. 2003;112:558–77.
 15. de Lange AH, Taris TW, Kompier MAJ, Houtman ILD, Bongers PM. Different mechanisms to explain the reversed effects of mental health on work characteristics. *Scand J Work Environ Health*. 2005;31(1):3–14.
 16. Zapf D, Dormann C, Frese M. Longitudinal studies in organizational stress research: a review of the literature with reference to methodological issues. *J Occup Health Psychol*. 1996;1:145–69.
 17. Lykken DT. What's wrong with psychology anyway? In: Chicchetti D, Grove W, editors. *Thinking clearly about psychology*. Minneapolis (MN): University of Minnesota Press; 1991. p 3–39.

Received for publication: 23 February 2006