HYPOTHESIS AND THEORY ARTICLE
published: 17 February 2011
doi: 10.3389/fpsyg.2011.00018

Methodological consequences of situation specificity: biases in assessments

Jean-Luc Patry*

Department of Education, University of Salzburg, Salzburg, Austria

Edited by:
Wei Pan, University of Cincinnati, USA

Reviewed by:
Wei Pan, University of Cincinnati, USA
Jason W. Osborne, North Carolina State University, USA

****Correspondence:***
Jean-Luc Patry, Department of Education, University of Salzburg, Akademiestrasse 26, A-5020 Salzburg, Austria.
e-mail: jean-luc.patry@sbg.ac.at

Social research is plagued by many biases. Most of them are due to situation specificity of social behavior and can be explained using a theory of situation specificity. The historical background of situation specificity in personality social psychology research is briefly sketched, then a theory of situation specificity is presented in detail, with as centerpiece the relationship between the behavior and its outcome which can be described as either "the more, the better" or "not too much and not too little." This theory is applied to reliability and validity of assessments in social research. The distinction between "maximum performance" and "typical performance" is shown to correspond to the two behavior-outcome relations. For maximum performance, issues of reliability and validity are much easier to be solved, whereas typical performance is sensitive to biases, as predicted by the theory. Finally, it is suggested that biases in social research are not just systematic error, but represent relevant features to be explained just as other behavior, and that the respective theories should be integrated into a theory system.

Keywords: bias, reliability, situation specificity, validity, typical performance, maximum performance, assessment, systematic error

INTRODUCTION

Research in social science is plagued by plenty of problems, some of the most important ones being methods biases. In its most liberal sense, a bias refers to systematic error in a research study (Gerhard, 2008); in the terms of Sackett (1979, p. 60) it is "(a)ny process at any stage of inference which tends to produce results that differ systematically from the truth." Since systematic errors are not obvious, theories for plausible rival hypotheses (Campbell, 1969) are necessary that provide explanations for the results of a study that differ from the one proposed by the hypothesis that is supposed to be tested.

According to Bungard and Bay (1982) researchers can react in four different ways to the risks of biases:

1. They can ignore biases; although this is the most frequent way, it is certainly the least appropriate one because the problem is not dealt with at all.
2. They can try to control for or even eliminate the potentially disturbing factors. A minority of methodically aware researchers follow this path. This is a possibility that has two disadvantages: (a) A systematic account of the potentially disturbing factors is necessary; this requires a theory that accounts for these factors, but such a theory is usually not used, mostly it is not even available. (b) The ecological validity (external validity with respect to situations outside of a research context) is in jeopardy because in "real life" the inhibiting factors that are eliminated or controlled for research purposes have a different impact than in the research study (limits of laboratory research, cf. Patry, 1982).
3. They can use alternative methods like the unobtrusive measures (Webb et al., 2000); this is done by even less researchers. This approach has clear limits: For many research questions one cannot find occasions for such assessments, and in studies using those that are available the internal validity is usually problematic (e.g., there is no random assignment to experimental and control groups).
4. They can question how the topic of interest has been addressed so far. Some researchers raise the question whether the problem of systematic error is due to the predominance of particular methods in research and to the lack of an adequate anthropological concept. In this interpretation, the bias problem reflects a restricted methodology. This approach is appropriate insofar it addresses the bias problem in its full consequence. Most of the researchers defending it, however, conclude that a completely different methodology is necessary (e.g., Guba and Lincoln, 2000); any other methodological approach, however, is just as much in need to address potential biases since it may be vulnerable to systematic errors although these might be different compared with the ones of the criticized methods.
5. A fifth way to react is proposed here: to develop theories that (1) account for the biases and that (2) can (and need to) be put in relation with the theory under investigation in a particular study. It is assumed that this way is the most appropriate because systematic errors are taken seriously and the respective threats – as far as they are known – are accounted for in the support and critique of a statement.

It is proposed that taking into account situations and the concept of action is a good base for this. The theory to be proposed in this paper, the Theory of Situation Specificity (TSS), addresses all these issues.

People often behave differently in different social situations, as has been mentioned repeatedly in the scientific literature, both in the older literature in which the topic was explicitly addressed (see the reviews by Mischel, 1968; Mischel and Peake, 1982; Patry, 1991a; etc.) as well as in newer studies (Ginsburg et al., 2006, to mention but one) although usually not in terms of situation specificity; and in many studies it is emphasized that situation specificity needs to be accounted for (among the many let me cite just Heppner et al., 2006; Radford, 2006; Sullivan-Marx, 2006).

Situation specificity, hence, seems to be a very important issue in psychology, in particular in social psychology, not to mention personality psychology where the discussion has started (e.g., Mischel, 1968; see below, Situation Specificity). Nevertheless, mostly situation specificity is just noticed within the study and not theoretically accounted for and not addressed as an important issue in current social research.

That situation specificity may have an immense influence on quantitative research methods has rarely been discussed, and even less accounted for theoretically, as would be necessary according to the fifth mode of dealing with biases. The aim of this paper is to address some of the relevant issues in methodology arising from situation specificity with particular emphasis on assessment. I will first give a brief introduction to the topic of situation specificity and present relevant elements of our TSS. In the subsequent sections central issues of reliability and validity are addressed and the TSS is applied. In the final section metatheoretical, theoretical, methodological, and practical conclusions are presented.

SITUATION SPECIFICITY

THE PROBLEM

In 1968, Walter Mischel published his book "personality and assessment," which triggered a broad debate. Mischel showed that in social behavior (particularly when assessed with systematic observation instead of questionnaires, cf. Peake, 1982), the correlations between assessments of behavior in different situations rarely exceed 0.30, while in cognitive psychology, consistency across situation is usually much higher. Looking back, Mischel (2004) comments the book as follows:

> Beginning with Hartshorne and May's (1928) studies of conscientiousness in schoolchildren, research had been driven by the assumption that the invariance of personality would be reflected in the stable rank-ordering of individuals in their behavior on any given dimension (e.g., conscientiousness, sociability, dependency), assessed with the cross-situational consistency coefficient. The assumption was rooted in a conceptualization of individual differences in social behaviors as direct reflections of behavioral dispositions, or traits. Dispositions and their behavioral expressions were assumed by definition to correspond directly, so that the more a person has a trait of conscientiousness, for example, the more conscientious the person's behavior was expected to be over many different kinds of situations, relative to other people. Given that assumption, the *persistent findings that the individual's behavior and rank order position on virtually any psychological dimension tends to vary considerably across diverse situations, typically yielding low correlations*, distressed the field and changed its agenda for years. (Mischel, 2004, p. 2; italics added)

Mischel was not the only researcher to notice this persistent finding: Much earlier (Heiss, 1948) and about at the same time (Bellows, 1963; Vernon, 1964; Hunt, 1965; Peterson, 1968; Wiggins, 1973; etc.) similar statements were published. But Mischel (1968) was undeniably the most influential publication. His results did not remain unchallenged, and there were several attempts to "solve the consistency problem," i.e., to develop methods which would yield higher correlations (e.g., Bem and Allen, 1974; Magnusson and Endler, 1977; Epstein, 1980; Buss and Craik, 1983; Snyder and Ickes, 1984; and many more; cf. also West, 1983; Kenrick and Funder, 1988; etc.); however, despite these authors' claims to have done so, the problem was not solved.

In most of these attempts, situational effects were considered as "error" (Mischel and Peake, 1983) or "noise" (Shoda, 2007, p. 327). Such an interpretation might be appropriate for some goals but not for others:

> Depending on one's purpose, the within-person variance – the interactional effects of persons with the conditions of their lives – may be as much part of the "true" fabric of human behavior, to be understood and analyzed, as is the abstracted categorization of a person's average performance in relation to a comparison group on the summary score of a more or less arbitrarily created test battery. (Mischel and Peake, 1983, p. 395)

I do not want to go into the debate that ensued (see Hoefert, 1982a; Schmitt, 1990; Moser, 1991; Patry, 1991a; Krahé, 1992; and others for details) but just mention that it has lost its vigor, and despite relevance as argued above, situation specificity is not an important topic in today's personality and social psychology research anymore, as can be seen in the abstracts of the 11th Annual Meeting of the Society for Personality and Social Psychology (SPSP, 2010), Among the 77 symposia and over 2000 posters, only few papers and posters deal more or less directly with situational issues and address the consistency debate as such (Cervone and Caldwell, 2010; Griffo and Colvin, 2010; Hensler and Wood, 2010; Sherman et al., 2010; Witt and Donnellan, 2010; this is a slight increase compared to the 2009 conference), some more deal with situational issues more or less directly or mention the impact of situations or situational factors without providing any deeper analysis of the situation dependency of behavior.

The result of the research on situation specificity can be summarized as follows (Patry, 2000):

1. In the social domain, i.e., when social behavior is at stake, situation specificity is the rule. For this there are exceptions under well-defined conditions (see for instance Price and Bouffard, 1974; Price and Blashfield, 1975). In the social domain, assessments using systematic observation by external observers have almost always yielded situation specificity; using questionnaires (e.g., personality questionnaires) however often does not show situation specificity unless the questionnaire is specifically conceived to be able to identify variation due to situations (e.g., in the tradition started with Endler et al., 1962).
2. In the cognitive domain, that is when cognitive abilities, intelligence, achievement, knowledge, and the like are at stake, cross-situational consistency is the rule, provided that in the two situations of interest the same ability is asked for. However, consistency seems to be lower than usually assumed; one can mention the problems in transfer (Salomon and Perkins, 1989; Detterman, 1993) or the problems of situated cognition (Brown et al., 1989; Cognition

and Technology Group at Vanderbilt, 1990; cf. also Lave, 1988; Greeno and the Middle School Mathematics Through Applications Project Group, 1998), and inert knowledge (Whitehead, 1967; Renkl, 1996): The subjects have learned the content but do not apply it in new situations although this would be appropriate.

3. One can distinguish roughly two types of research questions (Mischel and Peake, 1983): Question 1 deals with interpersonal variance, and situational variance is interpreted as measurement error ("measurement noise that obscure(s) a clear view of the person," Shoda, 2007, p. 327). This is the case for instance in differential psychology and in many studies in educational psychology. This is an important research approach, and it is not the aim of this paper to question its relevance, therefore the approach defended here cannot be called "situationist" (see also Mischel's, 2004, argumentation in this regard). However, it is also possible to address question 2 about how to account for *intra*personal variance, i.e., trying to account for the measurement error of question 1. This is what is done here.
4. For this it would be necessary to have a theory of situation specificity. However, only rudimental concepts of such a theory have been provided in the scientific literature so far. Among the most promising approaches one can mention:

   - person–situation interaction in the tradition of Magnusson and Endler (1977), which among others has been applied to personality (Endler, 1983) and anxiety (Endler, 1997);
   - the theory of intentional action (Ajzen, 1987);
   - Brunswick's lens model (Asendorpf, 1992);
   - Lewin's field theory (1951) and further developments thereof (e.g., Herber and Vásárhelyi, 2002);
   - Mischel's (1968, 1973) theory of social learning, of which the TSS is a further development.

5. All these theories have in common that situation specificity in the sense of an adaptation of the subject to the given conditions is appropriate in most situations. Someone is called "socially competent," for instance, if he or she has appropriate goals in a given situation, knows how to achieve them and acts accordingly. The goal structure and the means to achieve the goals, however, will vary from situation to situation.

DEFINITIONS OF "SITUATION" AND "SITUATION SPECIFICITY"

The definition of "situation" is much more complicated than usually assumed. The main problem is that a definition of situations used in a particular study must not be circular within the research design of that study (Thonhauser, 2007); a definition is circular if the situation is defined as conditions that yield behavioral differences, and then behavioral differences are "accounted for" by the situation. Such a circular definition must be avoided through an appropriate theory building and design (Patry, 2007). In addition, the definition of a situation must be in accordance with the hypothesis that is being tested in this study, which in turn will depend on the specific theory that is under investigation; hence taxonomies of situations for which universal applicability is claimed, as provided, for instance, by Bellows (1963), Eckes (1990), Frederiksen (1972), Hoefert (1982b), Kelley et al. (2003), and Price (1974), or concepts like the episodes (Barker and Wright, 1971), though inspiring, may be very questionable because of possible differences between the theory underlying the chosen taxonomy and the theory to be tested.

For the present purpose, I use Pervin's (1978) definition:

> A situation is defined by who is involved, including the possibility that the individual is alone, where the action is taking place, and the nature of the action or activities occurring. The situation is defined by the organization of these various components so that it takes on a gestalt quality, and if one of the components changes we consider the situation to have changed. (pp. 79f)

The components at stake in a particular study, then, will depend on the theory in such a way that circularity is avoided. Because of space restrictions, it will not be possible to elaborate in detail what this means in the specific studies or contexts that are discussed, although this could be done.

In TSS, it is assumed that situations have an impact on a subject's behavior only insofar as it is perceived by him or her, and with respect to the features (components in the terms of Pervin) that are perceived. Which of the situative features are relevant is determined by the subject's subjective theory, as far it is activated in the situation; hence it is the subject who determines which features are important and which are not.

Situation specificity means that relevant features of the behaviors of the *same* person[1] are different in different situations. This means that not the behavior as a whole is observed, but only certain characteristics thereof. For instance someone can behave cross-situationally consistently with respect to one behavioral feature (e.g., eye contact; or intelligent behavior) but situation specifically with respect to another (e.g., loudness of speaking; introverted behavior).

The opposite of situation specificity is cross-situational consistency. The higher the consistency (the lower situation specificity), the better one can predict the behavior feature(s) in situation S_2 given that one knows the one(s) in situation S_1. Three types of situation specificity can be distinguished (**Table 1**; from Patry, 2000, p. 16, with additions):

- *Relative* situation specificity: The *rank orders* of the subjects with respect to the interesting feature(s) of the behavior(s) are similar in both situations. This can be estimated using correlation coefficients: The lower the correlation, the lower the consistency, i.e., the higher the situation specificity. It is necessary to have at least two assessments of the same subjects in both situations. To establish a rank order it is necessary to have several subjects. It is not necessary to assess the same features in both situations; for instance, responding to a questionnaire (B_1 in S_1) can be correlated with observations of the behavior (B_2 in S_2), provided it is assumed that the features addressed by B_1 and B_2 belong to the same theoretical construct (e.g., a

[1] In special cases, different groups of people in different situations can be compared provided the groups can be assumed to be equal (usually through random assignment).

Table 1 | Types of consistency (with additions from Patry, 2000, p.16; translated from German).

	Relative consistency or situation specificity	Absolute consistency or situation specificity	Coherence
Definition	Rank order of subjects similar	Absolute value of (groups of) people equal	Reliable behavior patterns
Key figure	Correlation or the like	*t*-Test, ANOVA, ARIMA, etc.; possibly χ^2	ANOVA, Interaction P × S
Indicator of situation specificity	Small correlation (absolute value; low effect size)	Significant differences (high effect size)	High variance accounted for P × S
Subjects	Same	Same or similar (through random assignment)	Same
Number of subjects	Several	One or more	Several
Assessment tool	Similar or different	Similar	Same
Required scale level	At least ordinal	Nominal possible	Usually interval

questionnaire on introversion, B_1, and an observation of introversion, B_2). The scale level of the assessed features must be at least ordinal (requirement for correlations).

- *Absolute* situation specificity: The absolute values (or means) of the feature under investigation are different between situations. This can be tested using comparison of means [*t*-test with dependent samples, analysis of variance (ANOVA) with repeated measures] or differences in pairs of values (Wilcoxon); in case of nominal features tests like Chi-square can be used. It is possible to randomly assign the subjects to two groups (one for each situation) and to compare the means (*t*-test with independent samples, ANOVA between subjects; Mann–Whitney). Significant differences (or more appropriately, high effect sizes) represent situation specificity. It is possible to assess situation specificity for one single person, provided sufficient estimates of the behavior are available (e.g., time series with several observations in situation S_1 and several observations in S_2). The requirement for absolute consistency assessments is that the same feature is measured in both situations (or that the features can be transformed to be on the same scale, e.g., through appropriate *z* transformations).
- *Coherence*: Magnusson and Endler 1977), Endler et al. (1962) conceived the Person–Situation Interaction in terms of ANOVA: The higher the variance accounted for by the person × situation interaction, the higher the situation specificity. For this type of analysis the requirements are highest: The same subjects must be assessed with the same instrument in several situations, and the data must be on the interval level.

Comparing behavior in *similar* situations is an assessment of *stability*; this must not be confounded with cross-situational consistency. Mischel (1968) and many other studies do not question stability (relative consistency). Since stability is equivalent with retest reliability, or more generally with reliability according to any reliability estimates (see below, The Problem), and since reliability is not questioned in the literature, the problem of situation specificity as posited by Mischel (1968) and in the consistency debate (see above) cannot be reduced to a reliability problem, as has been done, for instance, by Epstein (1979, 1980); this has been underlined by Mischel and Peake (1982, 1983) and others. Rather, situation specificity is a substantial issue that requires a theoretical account for its own.

THE THEORY OF SITUATION SPECIFICITY

Based on Mischel (1968, chapters 6 and 7[2]) the TSS was developed and tested since the 1980s (details see Patry, 1989a, 1991a, 1992, 2000, etc.); those elements that are necessary for the further argumentation are described below. Since experience shows that misunderstandings are quite frequent in the discussions on situation specificity, the attempt is made to formulate as precisely as possible, at the risk of being overly detailed. The examples are taken from methodical issues and deal with the researchers' behaviors (in Reliability and Validity and following, the focus will be on the subjects' behavior).

Following Mischel (1968, chapter 6), his social learning reconceptualization of personality (1973), and its further development Cognitive Affective Personality System (CAPS, Mischel and Shoda, 1995; Mischel, 2007), the subject's goals and the means to achieve them are at the center of the theory. This is also in agreement with issue 5 in the review of the research on situation specificity presented above in Section "The Problem," as well as with Bungard and Bay's (1982) fourth and my own (fifth) answer to the bias problem (referring to actions) presented in the introduction. Accordingly, the central question is whether the means are appropriate to achieve the goals.

The basic assumption of TSS is that people tend to behave optimally with respect to achieving the respective goals, unless there are theoretically justified factors inhibiting this tendency. This optimization is done according to the subjective theory of the subject in the sense of Groeben et al. (1988). This means that behavior is assumed to be a rational action. Among these inhibiting factors one can mention emotions and arousal (as, for instance, in the Yerkes–Dodson-relationship).

People have often (but not always) several goals simultaneously (within one situation); the states of affairs being described by them can be incompatible (Rotter, 1954; Patry, 1997a, 2005a; Patry and Schrattbauer, 2000; etc.). This is particularly the case in social situations. Although one can assume that people have more than two

[2] These two chapters of "Personality and assessment" are rarely referred to in the literature on situation specificity. Even Mischel himself seems to have forgotten his own concepts: The theory he developed later on, for instance the statement "individuals are characterized by stable, distinctive, and highly meaningful patterns of variability in their actions, thoughts, and feelings across different types of situations. These *if . . . then* . . . situation–behavior relationships provide a kind of 'behavioral signature of personality' that identifies the individual and maps on to the impressions formed by observers about what they are like" (Mischel, 2004, p. 8) and other elements are less elaborate than the issues discussed in these two chapters of his 1968 book (Patry, 2009a).

incompatible goals, for sake of simplicity, the further discussion will deal with maximally two goals; in the case of more goals, the theory must be enhanced according to the same principles as discussed.

If people have different goals in different situations, one can speak of situation specificity *of goals*. Again, actually, it is necessary to focus on specific features of the goals; for instance, a student may have the general goal of finishing his or her studies in all study-related situations (lectures, group work, test): cross-situational consistency of the goals; on the level of the specific goals (understanding a concept, communicating, performing well) he or she will vary in function of the situation (situation specificity of goals).

One can distinguish two types of relationships between the means (behavior) and the outcome value[3]: "*The more, the better*" (relationship a in **Figure 1**) and "*Not too much and not too little*" (relationship c). Since any behavior feature B has a complementary feature non-B ("doing B" means "not doing non-B"; "doing more B" means "doing less non-B"), the third type, "The less, the better" (b in **Figure 1**) for B, can be regarded as "The more, the better" for non-B, hence a and b in **Figure 1** are not principally different and will therefore not be distinguished unless necessary.

Depending on the situation, the optimum of the relationships "Not too much and not too little" will differ (see **Figure 2**; Patry, 1991a, 2009b). This holds particularly if the goals are different in both situations. But it holds also in most situation pairs if the goal is the same (or highly similar) in both situations because it is quite frequent that in different situations different means are appropriate to achieve the same goal (so-called "phenotypic situation specificity"). Given the assumption that people tend to behave optimally, a relationship of the type "not too much and not too little" will yield full situation specificity (when different goals are at stake) and phenotypic situation specificity (with similar goals).

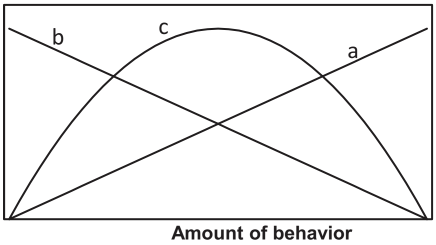

FIGURE 1 | Relationships between the behavior and its outcome value; a: the more, the better; b: the less, the better; c: not too much and not too little.

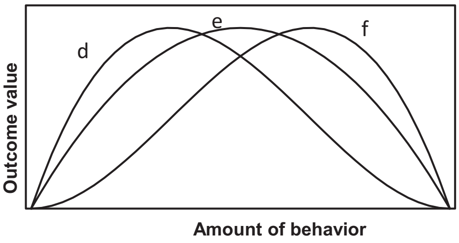

FIGURE 2 | "Not too much and not too little": different optima in different situations; d, e, f: different optima.

There are situations in which the subjects aim at performing their *best possible performance*. Then the principle "the more, the better" applies; the performance is limited by the subjects' ability. Complete consistency will then be the case if the following conditions are satisfied (cf. Sackett et al., 1988):

a. In each situation, all goals within the situation are compatible (no conflicting goals).
b. In both situations, the respective goals are to do the best.
c. In both situations, the "best" depends on the same ability.

An example is Datta's (1963) study on scientists' creativity. The correlation between the subjects' creativity test results (S_1) and the on-the-job ratings of creativity (S_2) was rho = 0.71 (relative consistency) provided that the subjects knew in S_1 that creativity was at stake. This means: In both situations the goals were the same, namely creative achievement, which addresses the same ability, and the subjects tried to do their best. In contrast, when the subjects were not told in S_1 that creativity was at stake (different goals in S_1 and S_2), the correlation was much lower (rho = 0.17). One can stipulate that the goal in S_1 was not to do the best with respect to creativity, actually the goal(s) the subjects pursued is or are not known; hence b and/or c in the above list were not satisfied. Subsequent research confirmed and differentiated this finding (see O'Hara and Sternberg, 2001; Chen et al., 2005).

These are the elements of the TSS that are important for the present purpose. More details and additional features of this theory are provided elsewhere.

Most of the theory's lawlike statements have been confirmed in many studies, many of which were not performed to test situation specificity hypotheses; quite the opposite: The typical published studies in this domain hypothesized that the behavior is consistent across situations, and this hypothesis was usually refuted unless both measures were with questionnaires (Patry, 1991a).

The TSS is a theoretical framework that can be used to integrate other theories. The TSS, by itself, is not sufficient because of the risk of a circular definition of situations. One relationship for theory integration is the superordinate/subordinate status: A superordinate theory is more general than the subordinate one, while the latter concretizes some of the issues that are left open by the superordinate one. Mischel (2007, p. 271) claims that CAPS is a superordinate theory[4], "a general framework that spells out a possible underlying structure"; it is superordinate for the TSS which addresses, among others, the relationship between expectations (variable 3 in CAPS; here called "means") and goals (values in CAPS, variable 4) and between abilities (CAPS variable 1) and expectations, etc. TSS, in turn, is superordinate to other theories, such as the theory of self-presentation (Christensen, 1981).

[3] This requires that both behavior and goal features are assessed at least on an ordinal scale level; in the case of nominal scale level, as in most expected utility models, the analysis will be different but this cannot be done here.

[4] Actually he calls it a "meta-theory," which is not appropriate since a meta-theory is a theory about a theory whereas the relationship between a superordinate theory and a subordinate theory is between theories that both are about the same kind of things, in this case behavior.

METHODOLOGICAL CHALLENGES WHEN ASSESSING SITUATION SPECIFICITY

In contrast to the traditional comparison *between* people, assessing *intrapersonal* variability poses a certain number of challenges. Some of them will be discussed here. The most important issue is the question of dependence of assessments since two or more measures with the same people are necessary.

A first type of dependence is in *ability testing*. Take the example that the same intelligence test (speeding tests) is responded by the same subject twice, one immediately after the other. One can imagine that the subject has learned from the first test taking and therefore will be faster the second time; in Campbell and Stanley's (1963) terminology this would be a pretest effect on the post-test that will jeopardize internal validity. Because of such effects, reliability of tests is assessed using parallel tests, i.e., tests that are similar but not equal; split half and internal consistency reliabilities are special cases of this. If the same test is used (stability), the second testing is performed a certain time (typically at least a few weeks) after the first one so that it can be assumed that the subject has forgotten about the features of the test.

When observing *social behavior*, however, the interdependencies are different. A prototype for such an assessment is Flanders (1970): The observation system requires coding teacher and student behavior on 1 of 10 categories every 5 s (time sampling). The categories of teacher behavior include, among others, "lecturing" and "asking questions," the students' categories are "student responses" and "student initiation." In such an assessment, serial dependencies may occur (Dumas, 1986): It is quite likely that the behavior at time t_n has an influence on the behavior on time t_{n+1}, 5 s later. For instance, lecturing, asking questions, responding to questions or initiating a discussion usually last longer than 5 s, hence the conditional probability of category x at t_{n+1}, given category x at t_n, is higher than the unconditional probability of category x. Further, the category "student response" occurring at t_{n+1} requires the category "asking questions" at t_n (or earlier) since it is defined as teacher initiated student statement. To identify such dependencies Flanders has conceived transition matrices, and his research show very high dependencies. The problem here is not remembering (which would be a test related variable and hence an assessment bias) but rather the behavior itself that has serial dependencies; this is the case independently of any observation.

To study situation specificity, we have used the lesson interruption method (Patry, 1997b, 2000): The teacher interrupts the lesson at a pre-decided time (e.g., before changing to a new topic), and the students answer a short questionnaire (typically we have used about 12 questions, but in some cases – e.g., Patry et al., 2000 – the questionnaire was longer) about their observations for the last 15 min. This way those who are most concerned by the lesson – the students – report their observation of what, in their view, has actually been going on in the given situation (the last 15 min). Once the students are accustomed with this method, they will respond very quickly; the experience shows that this observation does not disturb the course of the lesson. A typical result of such an assessment is given in **Figure 3**: The same teacher in a vocational school teaches five separate classes (101 through 106; class 104 did not finish the assessment) one lesson per week. The assessments were, among others, about the teacher's control (amount of direct teaching) with five items on a scale from 1 through 5 (for instance, "In this unit the teacher has given the students much – vs. little – personal freedom"; one refers to much freedom, five to little freedom, i.e., highly direct teaching). The studies using this method (e.g., Patry, 2000) have shown that the students answer quite homogenously (high interobserver agreement).

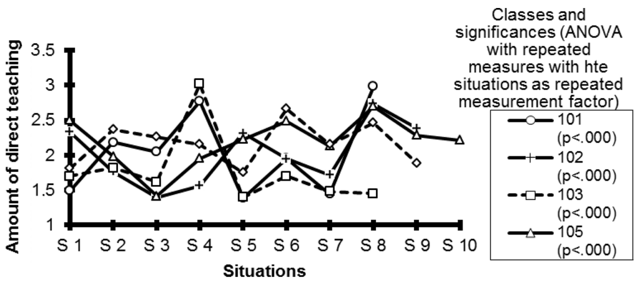

FIGURE 3 | Directivity of one teacher in five classes; teaching is once a week.

With this kind of data, the following steps are done:

1. Content validity: The items are judged for appropriateness with respect to the underlying theoretical construct (in the present case, for "amount of directive teaching").
2. Internal consistencies within the situation: Since several items (in the case mentioned above: five items) are used in the observation questionnaire by several students to assess the same construct, Cronbach's α can be calculated; if the items are homogeneous, internal consistency can become quite high compared to typical questionnaires (in Patry, 2000, the average of the situative internal consistencies for direct teaching was $\alpha_{\text{mean}} = 0.73$). In this case the mean of the items is calculated to form an average for each situation. These are given in **Figure 3**.
3. Autocorrelations (see, for instance, McCleary et al., 1980, p. 66ff) of the classroom averages per situation are calculated to check for serial dependencies. A serial dependency must be assumed if the autocorrelation coefficient for lag 1 (autocorrelation from t_n to t_{n+1}) is significant. In the above assessment with 1 week or more separating the assessments, no serial dependencies were found. Typically, even with shorter intervals, no such dependencies are found if the observation is in different situations.

4a. If no serial dependencies are found, we perform statistical analyses with the presupposition of serial independence. If the requirements like normal distributions are met (as was the case in this study) repeated measures ANOVAs are calculated per class with the situations as levels of the factor; the significances are reported in **Figure 3**.

4b. If the autocorrelations are significant, the assessments, cannot be considered as independent in the statistical sense. Classical statistical methods like repeated measures ANOVAs then cannot be used; instead methods like ARIMA (cf. McCleary and Hay, 1980) can be used if the conditions for this are satisfied.

5. It is also possible to use average values over situations. In this case it is necessary, first, to calculate the internal consistency with the situations as items and the means according to step 2 as values. The principle here is: Use the reliability (internal consistency) of the measure that you use.

Differences between situations (step 4a) often reach a high effect size; with each class represented in **Figure 3**, for instance, an ANOVA with repeated measure was calculated, and the results show highly significant effects, i.e., there are differences between the situations, and high variance accounted for (η^2). A teacher has once said that this method is like a seismograph of teaching. Interestingly, in contrast to these differences, the internal consistencies according to step 5 are usually very high. This contrast is due to the fact that the first difference (step 4a) refers to differences in the observed characteristics (e.g., the behavior of the teacher), while internal consistency refers to characteristics of the observers (in this case, the students).

RELIABILITY AND VALIDITY

Psychological research aims at achieving validity. Traditionally, a distinction between the validity of the assessment tool (Cronbach and Meehl, 1955; Campbell and Fiske, 1959) and the validity of the design (Campbell and Stanley, 1963) is made, and typically only one of them is discussed, isolated from the other. One could add to this the validity of the independent variable. Although issues of situation specificity are important for all three validities, because of space restrictions I will deal only with assessment.

THE PROBLEM

In their famous paper Campbell and Fiske (1959) state:

> Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods. (p. 83)

This can be seen as operational definitions of reliability and validity. The classical approaches to reliability, like internal consistency, split half reliability, parallel test reliability and retest reliability, are in full agreement with this operational definition; the same applies to approaches like interrater or intrarater reliabilities and the like. As to validity, Campbell and Fiske's convergent validity (which are the same as concurrent validity in the Cronbach and Meehl terminology) is a prototypical example for the application of their operational definition; in the multitrait–multimethod (MTMM) matrix, its informational value is enhanced by comparing it with the discriminant validities. The operational definition applies to other concepts of validity of assessment tools as well.

"Methods" in the above definition can be seen as a synonym of "situation" since the assessment method is one component of situation according to the definition (see Definitions of "Situation" and "Situation Specificity"). Substituting "method" with "situation" does not change the meaning of Campbell and Fiske's statements. Hence the definitions address directly the issues of stability (reliability) and situation specificity (validity). The low cross-situational consistencies reported by Mischel (1968), hence, mean that in social behavior criterion-related validity is quite low unless there are similar assessment tools; this indicates a strong instrument (or situation) bias.

Such instrument biases can be identified in the MTMM matrix: An instrument bias would result in a relatively high heterotrait monomethod (different constructs assessed with similar methods) correlation compared to the monotrait heteromethod correlation (convergent validity). And this is actually what is found in many MTMM analyses: Already Campbell and Fiske's synthetic example showed quite high correlations in the heterotrait monomethod triangles, almost the same size as the validity diagonals (convergent validities: monotrait heteromethod cells). In their examples taken from the literature (their tables 2 through 12), with few exceptions, the heterotrait monomethod correlations are at least as high as, if not higher than, the convergent validities. Indeed, they state in their summary:

> Measures of the same trait should correlate higher with each other than they do with measures of different traits involving separate measures. *Ideally, these validities should also be higher than correlations among different traits measured by the same method.*
>
> Illustrations from the literature show that these desirable conditions, as a set, are rarely met. Method or apparatus factors make very large contributions to psychological measurements. (p. 104; italics added)

These large contributions of method or apparatus factors have been found consistently in MTMM analyses of social behavior assessments since the introduction of the matrix by the authors of the MTMM technique (Campbell and O'Connell, 1967; and Fiske, 1982); later for instance Spector (1989) found no methods bias by just comparing correlations, but Williams et al. (1989), using confirmatory factor analysis, showed that the method bias was quite substantial. Many meta-analytical studies followed (e.g., Dickenson et al., 1986; Cote and Buckley, 1987; Marsh, 1990; Woehr and Arthur, 2003; Bowler and Woehr, 2006; etc.) in which consistently the method variance (in terms of TSS: variance accounted for by the situation) was substantial unless the methods were rather similar (which is in accordance with Mischel's, 1968, first conclusion presented above in The Problem). Obviously, there is a problem with the methods.

To address this question using the TSS it is necessary to distinguish what Cronbach (1970) has called "typical" and "maximum" performance.

TYPICAL AND MAXIMUM PERFORMANCE

For an analysis of assessment in terms of the TSS, first, it is necessary to check whether the behavior-outcome relationship is of the type "the more, the better" or "not too much and not too little." The former is related with cross-situational consistency, the latter with situation specificity. With respect to assessment, this distinction is the same as the one introduced by Cronbach (1970, p. 35ff). Tests to seek to measure maximum performance are used "when we wish to know how well the person can perform at his best" (p. 35), whereas typical performance refers to "what he is likely to do in a given situation or in a broad class of

Table 2 | Differences between maximum and typical performance assessments following Fiske and Butler (1963), Cronbach (1970), Wallace (1966), Willerman et al. (1976) and others (excerpt from Patry, 1991a, pp. 298–303).

Feature	Maximum performance	Typical performance
Assessed variable	Ability to respond	Disposition to respond
Generalization aimed at	What someone can do, but not what he or she will do	What someone is likely to do; the ability to do so is assumed to be given
Method	Usually assessed directly: The subject does, what the researcher is interested in	Assessed indirectly: The subject describes what he or she does or feels in certain situations or reacts to ambiguous material; sometimes observation is used
Instruction: What is assessed?	"This is an ability test." The subject is informed about the ability at stake (intelligence, knowledge, creativity, etc.)	Usually the subjects are not told that it is a (personality) test and about the disposition at stake to avoid reactive effects
Instruction: Right answer	"Give the right answer to each question!" "There is only one right answer"	"There is no right or wrong answer"
Instruction: How to answer	"Try to give your best!"	"Be as honest as possible!"
Instruction: Number of answers	"Do not expect to be able to answer all questions!"	"Please answer all questions, do not leave out any!"
Dealing with missing values	Missings are errors	Missings lead to elimination of the subject from the sample (or a guess what the answer would have been)
Instruction: Clearness	Instruction is not always clear, but clearness is aimed at	Instruction is not always clear, but ambiguity is often intended, particularly in projective tests
Implicit understanding	The subject assumes that the researcher wants him or her to do his or her best	The subject has no information about what the researcher aims at, he or she may guess (rightly or wrongly)
Relationship researcher-subject	Researcher controls the situation; for the test to be possible, the subject must accept his or her role; researcher and subject agree in their goals: harmonic relationship	Researcher controls the situation; for the test to be possible, the subject must accept his or her role; researcher and subject have different goals: relationship is not harmonic
"Difficulty" (probability of answers of a certain type)	True difficulty: There are items that the subject cannot answer (within the time restrictions); difficulty is important	All items can be answered in all ways by all subjects; "difficulty" plays no role
Robustness	Slightly differences in the formulation of the item and in context factors have no influence on difficulty	Slight differences in formulation of the items or context factors have an important influence on the results
"Upper limit"	There is an upper limit in performance: ability	There is no "upper limit"
Response strategy	Usually the strategy used by the subject is the one assumed by the researcher	Usually the researcher has no information about the answer strategy used by the subject
Consequences for the subject	Usually the subject knows quite well what consequences his or her answers will have	Usually the subject does not know how his or her answers will be interpreted and what consequences a specific answer will have (but he or she can guess)
Comparability	Assessments of different subjects are comparable: The test assesses the same for all subjects	Assessments of different subjects may assess different constructs (particularly in projective tests)
Reliability	Stable, high internal consistency	Lower stability, lower internal consistency; reduced applicability of test theory
Judgment criterion	The more, the better	There is an optimum: not too much and not too little; the optimum may differ from situation to situation

situations" (Cronbach, 1949). While Cronbach underlines that "(t)he classification scheme is a convenience in organizing our discussion, not a basis for theory" (Cronbach, 1949), I see it as an instance of the TSS.

A similar distinction has been provided by Fiske and Butler (1963); see also Willerman et al. (1976), Wallace (1966), Sackett et al. (1988, 2007) and others. Patry (1991a) provided distinctions between typical and maximum performance with respect to 28 issues; a short version is presented in **Table 2**, referring to paper–pencil or researcher-subject face-to-face assessments like tests, questionnaires, interviews, etc.

There have been some misunderstandings about the distinction between maximum and typical behavior. For instance Ackerman and Kanfer (2004) say that Hebb's distinction of Intelligence A and Intelligence B and Cattel's distinction of fluid intelligence (Gf) and crystallized intelligence (Gc) provide a reasonably close categorization of abilities that are associated with maximal performance and typical performance, respectively. (p. 121) The authors argue that the prototypical measures of what they call typical performance, WAIS-III, and Stanford–Binet IV, test "knowledge that the examinee has acquired and maintained over a long period of time." I agree that there is a difference between this type of intellect and the classical intelligence tests as assessed, for instance, with the Raven Progressive Matrices Test or the Culture Fair Intelligence Test (the authors' prototypes of maximal performance). Nevertheless, all these tests satisfy the conditions for maximum performance in the sense of Cronbach (1970): The issue in his distinction is not whether the ability at stake is basic or applied, but whether the assessment is a test of ability or of actual behavior and the like.

As an example for the two types of assessment, imagine an intelligence test as prototype of maximum performance and a questionnaire for extraversion as prototype of typical performance; both assessment tools are used in two contexts (situations): (S_1) a typical research situation with volunteers who know that the result will have no consequence on their life, and (S_2) a typical job application situation for a position as salesman. In both situations the subject will try to perform optimally: In S_1, since he or she is a volunteer, he or she will not be negativistic (in the sense of Weber and Cook, 1972) but try to impress the researcher (impression management, Christensen, 1981), while in S_2 he or she wants to get the job. For the intelligence test, the subject will try, in both situations, to perform at his or her best. For the extraversion questionnaire, the constellation is different: In S_1, the subject will answer as he or she thinks to be more or less appropriate, particularly depending on his or her self-concept according to his or her interpretation of the items (trying to follow the instruction to be honest, cf. **Table 2**). In S_2, however, he or she might identify the items as assessing introversion; since he or she thinks salesmen should be extraverted, he or she will tend to give the response assumed to represent extraversion. The presumptive employer (who, in this regard, has the same role in S_2 as the researcher described in the table and in S_1) is not interested whether the applicant will behave intelligently or in an extraverted way in all situations but what he or she will do in practical sales situations (S_3): For intelligence as well as for extraversion, predictions for future behavior are intended. The prediction from S_2 to S_3 will be much more valid with respect to intelligence (maximum performance) than for extraversion given all the constraints of typical assessments. The comparison of the two columns concerning maximum and typical behavior reveals other differences related with TSS (which cannot be accounted for in detail here) which show that the validity of the former is much higher than the validity of the latter.

The question is whether the issues addressed in the table are relevant in situations other than paper–pencil or face-to-face assessments, such as observations in natural settings or unobtrusive measures (Webb et al., 2000). A prototypical example is school: School grades are some kind of unobtrusive measures because the tests they are based on are not done for research purposes. For instance, mathematics tests, if well done, satisfy most of the conditions for maximum behavior mentioned in **Table 2**, with the teacher taking the role of the researcher. Other school tests may fit these conditions to a lesser degree; particularly the criterion "Instruction: What is assessed" is not always met: In many tests the ability at stake (or the criterion for right answers) is not told to the students, and maybe several criteria are relevant simultaneously, such as spelling, grammar, style, structure, and content in an essay; as to the content of the essay, the student might not know (or guess) what is important to the teacher and what is not. Since school tests aim at assessing the ability, the maximum performance column in **Table 2** can be interpreted as a checklist for good school tests.

In school tests there is the possibility that a student does not aim at performing to his or her best. One can imagine a student who does not want to be seen as a know-it-all by his or her peers or aims at a lower achievement for other reasons; in Weber and Cook's (1972) terminology this would be called a (partly) negativistic subject. In this case the principle "not too much and not too little" applies: The student shows a lower performance than would correspond to his or her ability, but not too low so that it does not become obvious that he or she does not do his or her best.

Maximum performance conditions are very rare. Besides school, maximum performance can be found in sports (e.g., in a 100-m dash, the athletes run as fast as they can: "the more, the better") and, to some degree, in job situations (e.g., Sackett, 2007). It is quite straining to perform at one's best in some regard over an extended period of time. We follow the principle "the more, the better" only upon request, either when asked to do so (school, job, sports) or when the circumstances are such that one has a goal that pushes one to go at one's limits (e.g., when trying to catch a bus, one runs as fast as one can), but in our daily life maximum performance is the exception and not the rule; with respect to walking or running speed, mostly it is typical performance ("not too much and not too little"), and it depends on the situation (e.g., whether I am in a hurry or I have time, etc.).

MAXIMUM PERFORMANCE AND SITUATION SPECIFICITY

In terms of the TSS, situations in which maximum performance is asked for are such that the subject has only one goal: to perform at his or her best, i.e., to get at the limit of his or her ability; similarly, Cronbach (1970) refers to such tests as "tests of ability." Test conceivers and users are required to design and apply assessment tools in such a way that the subjects have no goals to perform lower than their ability level, and they have to provide a situation in which there are as little obstacles to this as possible.

The convergent, concurrent and predictive validities will be applicable if the conditions for maximum performance mentioned above are met for both tests (the test to be validated and the criterion test). This is the case if

1. the subject knows the required performance and the criterion to judge its quality;
2. the criterion is such that the more of the behavior, the better;
3. the goal of the subject is to perform as well as possible;
4. the subject has no simultaneous incompatible goal (i.e., the goal of not doing his or her best) in this situation; and
5. there is no other factor that inhibit the maximal effort.

The best example for this is intelligence which can be considered as the best predictor of school performance (e.g., Bratko et al., 2006; Spinath et al., 2006; Freudenthaler et al., 2008); it seems that at least the first four conditions are satisfied: The subjects know what is expected from them, the relationship is "the more, the better," the students try to perform at their best, and there are few incompatible goals if any. Whether the fifth condition applies in school might be questioned in some cases. For instance personality variable can increase the variance accounted for: conscientiousness (Bratko et al., 2006), self-esteem (Freudenthaler et al., 2008), and ability self-perceptions and intrinsic values (Spinath et al., 2006), etc.

The impact of the respective variables is consistent with the five conditions:

- Conscientiousness is "a trait referring to individuals' level of dutifulness, achievement striving, and organization. Importance of Conscientiousness in educational settings is self-explanatory: Being organized, disciplined, and motivated to succeed has no doubt beneficial effects on students' study habits, affecting their level of effort and commitment with the course." (Bratko et al., 2006, p. 132) Conscientiousness, hence, is linked with conditions 3 (perform as well as possible) and 4 (it means that the goal of performing as well as possible is the students' dominant purpose). Lack of conscientiousness would therefore be an inhibiting factor.
- Self-esteem, ability self-perception, and intrinsic values relate with the confidence to perform well and hence with the effort put into performing well; inversely, low self-esteem, low ability self-perception, and low intrinsic values can be seen as inhibiting factors (see condition 5).

Other variables accounting for school performance, like school-related intrinsic motivation, school anxiety, and performance-avoidance goals (only for boys) and work avoidance (only for girls; Freudenthaler et al., 2008) or differential influences of the variables in different content areas (Spinath et al., 2006) confirm the importance of the five conditions but show also that they are just a framework; for each of them specifications are necessary with relation to the specific conditions.

In many laboratory research designs, for instance in cognitive psychology, maximum performance tools are used for the assessment of competence. A classic example is Ebbinghaus' (1913) use of nonsense syllables: The syllables were learned according to a specific paradigm until they could be fully reproduced. The five conditions are satisfied: The more syllables the subject can reproduce, the better the performance; if all syllables can be reproduced, the ability (in this case, knowledge) is at its peak. Ebbinghaus tested what strategy would lead the learner the quickest to this performance. The advantage of this procedure, with respect to the issues discussed here, is that this way the assessments can be performed with high reliability and, if done appropriately, with high validity. Similar approaches have been used in most studies on memory and for other cognitive tasks with great success. In my view this is one of the most important reasons for the rapid progress in cognitive psychology.

TYPICAL PERFORMANCE AND SITUATION SPECIFICITY

In contrast to cognitive psychology, the success in social and personality psychology has been much slower. In the 1970s and 1980s there was even a talk about social psychology's crisis (e.g., Elms, 1975). I pretend that the problems with assessment are the most important source of difficulties and that this is due to the "not too much and not too little" -relationship (or typical performance) discussed above.

It is striking that in social psychology deception (Milgram, 1963, is a classic example for this) with the risk of suspicion (McGuire, 1969) has been used extremely often. According to Hertwig and Ortmann (2008, p. 65), 50% and more of the articles published in the Journal of Experimental Social Psychology and in the Journal of Personality and Social Psychology employed deception. Other methods like projective tests (e.g., the Rorschach test), unobtrusive assessments (Webb et al., 2000), subtle (instead of obvious) items (Lanyon, 1984)[5], naturalistic observation situations (e.g., "waiting situation," Mehrabian, 1971) have been used; since many of them are ethically questionable, there must be strong reasons not to avoid them.

The aim is to reduce some of the many potential biases (Sackett, 1979, cataloged "35 biases that arise in sampling and measurement," p. 51) that are linked with features like social desirability (Edwards, 1957), reactance (Brehm, 1966), demand characteristics (Orne, 1969), subjects' motives (Weber and Cook, 1972), experimenter effects (Rosenthal, 1976), faking (Pauls and Crost, 2005), etc. This can be seen as a means to eliminate potentially disturbing factors (see the second reaction according to Bungard and Bay, 1982, presented in the introduction), the disturbing factor being the subjects' knowledge about relevant issues. One cannot imagine, for instance, that Milgram could have done his studies with the subjects knowing that the real theme of the study was not learning but obedience. Interestingly, Geller (1978) found that involvement, i.e., being absorbed by the situation, forgetting everything else, etc. (this is comparable with flow in the sense of Csikszentmihalyi and Csikszentmihalyi, 1988) yielded similar results as Milgram even when the subjects knew the aim of the study beforehand – but they had forgotten it; according to the TSS, the subjective theory (the goal structure) is only relevant insofar as it is activated in the particular situation, and involvement in the sense of Geller leads the subject to take into account the goals and means he or she would do spontaneously (without influence by the researcher).

The conditions for convergent or criterion validities in social research situations according to the TSS are much more complex than for maximum performance. Given that people tend to behave optimally as assumed in the TSS, the following issues must be taken into regard:

1. What are the goals in the respective situations? Are the goals similar, or are they different?
2. What are the means that are appropriate to achieve the goals according to the subject?

Let us look at the goals first. In social situations, according to the TSS, one can assume polytely (multiple goals) with conflicts on the goal and on the means levels and – typically – compromises. One can further assume that the respective goals depend on the situation, which is determined by how the subject perceives it, particularly by the features seen as important. And the most important

[5] Lanyon (1984, p. 674): "The question of the validity of subtle vs. obvious items continues to be debated. The notion that subtle items make a small but unique contribution to valid variance (…) has been eroded by the literature.", and he cites several studies that "showed that scales using subtle items were less valid than those composed of obvious items" (Lanyon, 1984).

features in order to behave optimally are those that determine the goals the subject wants to pursue. Hence the first question is how the goals are influenced in the research situation.

The theoretical accounts of biases presented above unanimously emphasize goals related with the awareness of the fact that one is being observed. This is obviously the case for social desirability (Edwards, 1957), the aim being to present oneself in a positive way, and with the subjects' motives according to Weber and Cook (1972) and faking. Reactance (Brehm, 1966) is a special case: The subject perceives restrictions imposed upon him or her through the research situation (specific interpretation of the situation) and aims at re-establishing the freedom. One can further assume that the experimenter effect (Rosenthal, 1976) is closely related to goals because the subject wants to respond positively to the perceived demands of the respective situations, and the same applies to demand characteristics (Orne, 1969). Even the use of naturalistic observation situations is related with the subjects' goals, as has been shown, for instance, by Higgins et al. (1983): The aims of the subjects are different depending on whether they know that they are being observed or not.

In most examples from Webb et al. (2000) the advantage of unobtrusive methods is that the subjects' goals are not influenced by the fact that a research is taking place. Subtle items have the function to hide the real goal of the research situation so that the subjects' goals to behave according to the perceived requirements of the situation do not interfere with the issues addressed in the hypothesis, and deception has the same justification. Reactivity of behavioral observations (e.g., Christensen and Hazzard, 1983) can also be accounted for through goal shifts.

Overall, most of the biases are due – at least partly – to additional goals because the subjects know that they are observed for research purposes as opposed to natural or naturalistic situations in which the subjects are unaware of the observation. According to Christensen (1981), the main goal is self-presentation; while Christensen focuses on the motives discussed by Weber and Cook (1972), the argumentation above suggests that this kind of goals may be as important in most other sources for biases as well.

However, the effects are not always the same, as can be seen when the same bias is analyzed repeatedly. The effect sizes of the impacts of knowing about being observed often cannot be replicated, and small changes in the research conditions may completely alter the size of the bias or even eliminate it at all. With respect to reactivity effects, for instance, Jacob et al. (1994) state that various reviews of the literature on the observation of distressed and non-distressed families "have concluded that findings have been diverse and inconclusive" (p. 355), and they attribute this status "to the scarcity of methodologically sound studies on this topic" and "to the fact that many variables can affect the strength and direction of reactivity. Furthermore, the theoretical frameworks relevant to explaining reactivity effects (…) have been extremely limited." (Jacob et al., 1994) While the first reason (few methodologically sound studies) may be true, the second and third reasons – which are related since it would be necessary to integrate the variables in a theoretical framework – seem far more important; however, although the authors call for a theory, they do not provide one except for saying that the most important factor was observer salience, or obtrusiveness. They report that some studies suggest that mothers try to maximize positive behavior, while other studies did not replicate this finding. In their own study, "results indicated relatively few instances of reactivity effects (…). The larger literature on reactivity effects in family research (…) yields the same general conclusion." (p. 360) In this context it is not so important whether reactivity effects occur or not; the relevant issue is that *it is possible* that they occur, and if no theory is available to predict such effects, researchers must always live with the risk of reactivity.

TSS can account for these biases as follows:

- If the subjects know that they are being observed for research purposes, they will have the goal of self-presentation (Christensen, 1981).
- Given polytely, the subjects have other goals besides self-presentation.
- If a situation has been set up purely for research reasons (typically laboratory research), i.e., if the decisions made in the situation by the subjects have no impact on the subjects' further life or on other people involved (except for the researchers) one can assume that the subjects have few other goals if any. But even in this case the specific form of self-presentation will depend on a multitude of factors that cannot be discussed here. This can be a question of goals and/or a question of means.
- If the outcomes of the subject's behavior are of any future importance outside of the research context, corresponding goals will become important. For instance, a teacher in the classroom may be aware of observers and may try to please them, but at the same time he or she will have the practical goals which include conveying the content as intended, keeping discipline, engaging the students in learning, etc. (e.g., Hofer, 1984; Krampen, 1984), hence he or she cannot concentrate uniquely on impression management on the observer.
- In both cases, the goals and/or the corresponding means will probably be incompatible to some degree (goals and means conflicts).

An adaptation of this concept has been presented by Patry (2004). It is not possible here to go into further details. It must be mentioned, though, that the factors playing a role in typical performance are much too complex to permit at the current state of the art to predict with reasonable precision what biases will have an influence – one can just say that the likelihood of biases is very high. The framework presented above provides at least a series of relevant factor groups and a first approach for theory. It will be important to apply this framework to different studies to develop it further.

CONCLUSIONS

The distinction between typical and maximum behavior is closely linked to the distinction between "not too much and not too little" and "the more, the better" in the TSS. The impact on research is considerable. The research results are in agreement with these theoretical assumptions. For instance Follman (1984) reports the data presented in **Table 3**. As can be seen, maximum performance assessments in both situations (here: assessment instruments) yield high correlations, provided the ability at stake is the same in both

Table 3 | Real-world correlations (excerpt from Follman, 1984, p. 702; the references are omitted).

Variables	*r*
IQ test reliability	0.90s
Standardized school achievement test reliabilities	0.90s
IQ and school achievement–grade 1	0.85-0.90
IQ and school achievement–college from high school	0.50-0.55
GRE and graduate school grade point average	0.00-0.40
IQ and memory (higher with age into adulthood)	0.50-0.70
THE UBIQUITOUS 0.35 CORRELATION	
School achievement (cognitive) and affective	0.35
School achievement and self-concept	0.35
School achievement and motivation	0.35
School achievement and student ratings of teacher effectiveness	0.44
IQ and self-concept	0.35
IQ and creativity	0.35

assessments (this is not the case, for instance, when correlating IQ with creativity, last row in the table). Whenever at least one of the variables is typical performance, the variance accounted for (square of the correlation) is about 10%.

Studies comparing typical and maximum performance confirm the hypotheses that emanate from the theory as described; for instance, most studies reported in the articles of the special issue of Human Performance (Klehe et al., 2007) dealing with job performance are in agreement with the theory (Klehe et al., 2007; Mangos et al., 2007; Marcus et al., 2007; Ones and Viswesvaran, 2007), others provide results that go beyond the theory and may give hints as to how to improve it (Smith-Jentsch, 2007). In any case the concept proved to be quite promising.

DISCUSSION

Issues of situation specificity have been neglected in psychology in general and in methodological discussions in particular. However, such issues are addressed frequently with respect to methodical problems and in practical contexts, yet without being analyzed systematically: Often it is just said that a specific phenomenon must be regarded as situation specific, but there is no theoretical account. Actually, whenever methodical or methodological problems are addressed in a publication – particularly when social behavior is at stake –, it is likely that situation specificity is relevant in some way.

META-THEORETICAL ISSUES

Given the complexity of the theoretical framework that should take into account all the issues mentioned above – situations, multiple theories, and many more – it becomes impossible to conceive studies that address it as a whole. Rather, it is necessary to follow several research strategies.

First of all, research *programs* (see, for instance, Herrmann, 1976) instead of single studies are appropriate; each study within this program contributes a piece to the full picture. We have followed such a program for situation specificity for the last 30 years (Patry, 2005b).

Secondly, it is important to replace the traditional "theory testing approach" by a more modest ambition. The best we can do is what Dewey (quoted from Phillips and Burbules, 2000, p. 31) has called "*warranted assertibility*":

> When knowledge is taken as a general abstract term related to inquiry in the abstract, it means "warranted assertibility." The use of a term that designates a potentiality rather than an actuality involves recognition that all special conclusions of special inquiries are parts of an enterprise that is continually renewed, or is a going concern.

A warranty, in this context, is a support of a statement or its credibility (Phillips, 1997). Every statement that is claimed to be scientific needs to be backed up in a reasonable way. The more arguments in favor of a statement are provided, and the more substantial they are, the more credible the statement is (see also Phillips and Burbules, 2000, p. 3).

THEORETICAL ISSUES

Biases are usually interpreted as error, as distortion of the data, as something that is unwelcome and must be avoided. However, these biases may be "as much part of the 'true' fabric of human behavior, to be understood and analyzed,"[6] as the behavior that is addressed in the theory and in the hypotheses. Why should the biases not be some kind of truth, why should it be completely different from everything else? Instead it is much more appropriate to assume that there is not a contradiction between the behavior one is interested in and the behavior associated with biases.

For this it is necessary to combine the theories used – the theory under investigation ("research theory") and the theory that accounts for the biases. Whether the TSS is an appropriate framework for this endeavor will depend on the research question and on the research theory. Not all research theories are fully compatible with the TSS. The claim here is not that the TSS is the only approach that accounts for biases; instead it is a proposition, but if it is not compatible with the research theory other theories of bias must be used.

How can this integration of theories be performed? An example is research in education. By definition and in practice, education means to pursue many goals simultaneously (polytely); many of these goals are incompatible: Teachers have educational goals like the students' mastery and performance (e.g., Wolters and Daugherty, 2007; Darnon et al., 2010), fostering the students' social competence and emotional development and other social goals (Allody, 2010), keeping discipline (Reyna and Weiner, 2001) etc., but they have also personal goals like emotional regulation (Sutton, 2004) etc. Publications on didactics usually underline the antagonist structure of teaching; Becker (1984), for instance, lists more than 20 tradeoff pairs, such as "foster interaction between students – and support individual work" or "respond to students' questions – and leave some questions unanswered."

When observing teaching overtly for research purposes, it is necessary according to the TSS (i) to take into account all presumptive goals of the teachers have typically (or, if available, all goals they actually have) in such teaching situations, (ii) to check whether

[6] This is a quote from Mischel and Peake (1983, p. 395; cf. above, The Problem) but applied to a slightly different context.

and how far these goals are being influenced by the presence of an external observer, (iii) to ask whether the presence of the observer triggers new goals, and (iv) to estimate the balance, i.e., the relative weights of the different goals in this situation. Depending on the situation, the teacher's goal balances will be different:

S_1 The teacher is alone with his or her students; this is the regular teaching situation. Variations in the teacher's goals are likely: The balances in the sense of Becker may differ from situation to situation.

S_2 A researcher has asked the teacher to implement a new teaching method and is recording the teaching by video. The teacher will probably aim at implementing the new teaching method to the best of his or her knowledge at the expense of other goals; if he or she does not agree with the new method, he or she will sabotage it (negativistic subject in the sense of Weber and Cook, 1972).

S_3 The parent of a student sits in the back of the classroom. The focus of the attention on this specific student will probably have a high priority within the teacher's objectives.

S_4 The principal of the school observes the teacher. The teacher is likely to aim at showing that he or she is a good teacher according to the principal's values (which may be different from the teacher's).

S_5 In the case of an internship, the supervisor judges the student teacher's actions. The student teacher's most important goal will probably be to comply with the preferences of the supervisor (Arnold et al., 2011).

In all these cases the teacher's goals and behavior are likely to be different from the ones in S_1. It might well be that during the course of teaching, the teacher forgets the fact that he or she is being observed (e.g., flow or habituation) – then the impact of the presence of the observer in any of the situation will be reduced, and the behavior resembles more the one in S_1.

Each of the goals is linked with a theory that describes how to achieve the goal (and which then is subordinate to the theory of situation specificity as discussed in The Theory of Situation Specificity): To account for the teacher's teaching of subject matters (S_1), a didactical theory may be appropriate (e.g., "this is constructivist teaching"), while for the other situations, a theory of self-presentation might be used (for S2, for instance, see Christensen, 1981) which in turn refer to still other theories (e.g., in the case of S_5, to the supervisor's favorite didactical theory). It might also be necessary to distinguish scientific theories that can describe and explain the teacher's behavior (e.g., an expected utility model, Feather, 1988) and the teacher's subjective theories (theories dealing with questions like "How can I reach the goals that I want to achieve?" and "How can I balance the different goals?" Patry, 2005a).

This example also illustrates another consequence of applying the TSS to research situations. The question is whether it is possible to make predictions about the behavior of the teacher's in S_1 based on an observation in one of the situations S_2 through S_5. This is the question of generalizability of the findings, or in other words of the domain of validity of the scientific statement that emanates from the research (Patry, 1991b). One can attempt to make such predictions without taking into account the TSS (reaction 1 in the sense of Bungard and Bay, 1982); however, the validity of such a statement is highly questionable. Including the situational dependency of the goals (and hence of the means to achieve them) in the theoretical framework may at least address the question of generalizability; it might well be that it can then lead to more precise predictions. But until this can be done, much more research will be necessary. Nevertheless, this incertitude is still much more valuable than falsely pretending that there is no problem at all.

METHODOLOGICAL ISSUES

Warranted assertibility (see Meta-theoretical Issues) can be improved by following the principles of critical multiplism (Cook, 1985; Patry, 1989b; Shadish et al., 2002). Hetherington (1997) distinguished two dimensions of research: thoughtless vs. thoughtful research and single vs. multiple methods; critical multiplism refers to thoughtful multiplism, which means systematic, rational theory-driven multiplism with the researcher being well aware of the problems and biases. The attempt is to compensate the biases a given method has or may have by using a different method that has different biases. This permits, if not to correct for biases, to identify whether such biases are present. Campbell and Fiske's (1959) multitrait–multimethod matrix that was discussed above is a prototype of such an approach (for details, see Patry, 2008).

PRACTICAL CONSEQUENCES

Practitioners (e.g., teachers, social workers, parents) have always known that social behavior is situation specific; in particular they have deliberately treated different children differently. For social scientists, it is different: Although research has shown very early that social behavior is situation specific (the first studies addressing hypotheses of situation specificity date from the late 1920s, e.g., Newcomb, 1929) they have neglected it. Only within personality psychology a debate about cross-situational consistency has erupted after Mischel's "Personality and assessment" (1968), but the focus of this debate was on retaining or abolishing the concept of personality. This question falls short of the importance of situation specificity in social research, as I have tried to show in this paper: Situation specificity and its impact on biases cannot be argued away but must be recognized; denying this would be following the first strategy Bungard and Bay (1982) mentioned: acting as if there were no biases.

One can guess why the topic of situation specificity has been so much neglected. There are several plausible reasons:

- When attributing reasons for the behavior of *other* people humans tend to use dispositional theories, i.e., theories that do not take into consideration situations; when attributing one's own behavior, however, people tend to refer to concepts like goals and means to achieve them in situations (Jones and Nisbett, 1971). It might well be that social scientists focus on other people and not on themselves. Some researchers underlined the role of their intuition, most drastically Bem and Allen (1974) with the following statement:

> We are not here denying the well-documented biases and illusions which plague our intuitions, nor do we claim that the more formalized idiographic procedures used by clinicians have a better track

> record in terms of predictive utility than nomothetic ones; they do not (Mischel, 1968). But in terms of the underlying logic and fidelity to reality, *we believe that our intuitions are right; the research, wrong.* (Bem and Allen, 1974, p. 510; italics added)

Such a statement is surprising for recognized scientists and shows how strong this intuition seems to have been. It must be mentioned, though, that the opposite intuition is very frequent. When submitting papers on situation specificity I got quite contrasting reviews; some reviewers rejected a paper with the argument that there is no situation specificity, while others rejected the very same paper with the argument that situation specificity is trivial. I hope I could show that neither argument is tenable.

- Another possible reason for neglecting situation specificity is that acknowledging it would threaten the own research. This was the case already in very early empirical research studies: about 1890 James McKeen Cattell developed "standard series of tests to be applied for 'discovering the constancy of mental processes, their interdependence, and their variation under different circumstances'" (Watson, 1959, S. 3) However, when the results showed situation specificity (low correlation between the measures and external criteria), the psychologists soon lost interest in the topic (Watson, 1959, p. 4), instead of investigating the reasons for the surprising results. Watson refers here to concurrent validity, and any threat to validity puts the research in jeopardy. I have shown elsewhere (Patry, 1991b) that situation specificity also threatens external validity, i.e., the generalizability of the results of a study. Many more challenges have been mentioned above. One can assume that researchers do not like to question their research and hence prefer to deny or ignore the problem.
- Accounting for situation specificity renders research very complicated on all three levels (meta-theories, theories, and methods, see above, Meta-Theoretical Issues through Methodological Issues), and few researchers want to render their research more complicated than necessary.

These may be explanations, but not excuses. Valid research does not permit ignoring or denying recognized threats to validity, rather these threats need to be accounted for and dealt with appropriately. I hope that the present paper is a contribution to this.

REFERENCES

Ackerman, P. L., and Kanfer, R. (2004). "Cognitive, affective, and cognitive aspects of adult intellect within a typical and maximal performance framework," in *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development*, eds D. Y. Dai and R. J. Sternberg (Mahwah, NJ: Erlbaum), 119–141.

Ajzen, I. (1987). "Attitudes, traits, and actions: dispositional prediction of behavior in personality and social psychology," in *Advances in Experimental Social Psychology*, Vol. 20, ed. L. Berkowitz (New York: Academic Press), 1–63.

Allody, M. W. (2010). Goals and values in school: a model developed for describing, evaluating and changing the social climate of learning environments. *Soc. Psychol. Edu. Int. J.* 13, 207–235.

Arnold, K.-H., Hascher, T., Messner, R., Niggli, A., Patry, J.-L., and Rahm, S. (2011). *Empowerment durch Schulpraktika: Perspektivenwechsel in der Lehrerbildung.* Bad Heilbrunn: Klinkhardt.

Asendorpf, J.-B. (1992). A Brunswikean approach to trait continuity: application to shyness. *J. Pers.* 60, 53–77.

Barker, R. G., and Wright, H. F. (1971). *Midwest and its Children. The Psychological Ecology of an American Town*, 2nd Edn. Hamden, CT: Archon.

Becker, G. E. (1984). *Durchführung von Unterricht. Handlungsorientierte Didaktik, Teil II.* Weinheim: Beltz.

Bellows, R. (1963). "Toward a taxonomy of social situations," in *Stimulus Determinants of Behavior*, ed. S. B. Sells (New York: Ronald), 197–212.

Bem, D. J., and Allen, A. (1974). Predicting some of the people some of the time. The search for cross-situational consistencies in behavior. *Psychol. Rev.* 81, 506–520.

Bowler, M. C., and Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *J. Appl. Psychol.* 91, 1114–1124.

Bratko, D., Chamorro-Premuzic, T., and Saks, Z. (2006). Personality and school performance: incremental validity of self- and peer-ratings over intelligence. *Pers. Individ. Dif.* 41, 131–142.

Brehm, J. W. (1966). *A Theory of Psychological Reactance.* New York: Academic Press.

Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educ. Res.* 118, 32–42.

Bungard, W., and Bay, R. (1982). "Feldexperimente in der Sozialpsychologie," in *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen*, ed. J.-L. Patry (Bern: Huber), 183–205.

Buss, D. M., and Craik, K. H. (1983). The act frequency approach to personality. *Psychol. Rev.* 90, 105–126.

Campbell, D. T. (1969). "Prospective: artifact and control," in *Artifact in Behavior Research*, eds R. Rosenthal and R. Rosnow (New York: Academic Press), 351–382.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105.

Campbell, D. T., and O'Connell, E. J. (1967). Methods factors in multitrait-multimethod matrices: multiplicative rather than additive? *Multivariate Behav. Res.* 2, 409–426.

Campbell, D. T., and Stanley, J. C. (1963). "Experimental and quasi-experimental designs for research on teaching," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally), 171–246.

Cervone, D., and Caldwell, T. L. (2010). Explaining personality coherence from the bottom up: applying the Kapa model of personality architecture. *Paper read at the 11th Annual Meeting of the Society for Personality and Social Psychology. In SPSP: The 11th Annual Meeting of the Society for Personality and Social Psychology* (p. 90). Available at: http://www.spspmeeting.org/documents/SPSP2010_Program.pdf on [Accessed on 23 January 2010].

Chen, C., Kasof, J., Himsel, A., Dmitrieva, J., Dong, Q., and Xue, G. (2005). Effects of explicit instruction to "be creative" across domains and cultures. *J. Creat. Behav.* 39, 89–110.

Christensen, A., and Hazzard, A. (1983). Reactive effects during naturalistic observations of families. *Behav. Assess.* 5, 349–362.

Christensen, L. (1981). Positive self-presentation: a parsimonious explanation of subject motives. *Psychol. Rec.* 31, 553–571.

Cognition, and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educ. Res.* 19, 6, 2–10.

Cook, T. D. (1985). "Post-positivist critical multiplism," in *Social Science and Social Policy*, eds R. L. Shotland and M. M. Mark (Beverly Hills, CA: Sage), 21–62.

Cote, J. A., and Buckley, M. R. (1987). Estimating trait, method, and error variance: generalizing across 70 cross-validation studies. *J. Mark. Res.* 24, 315–318.

Cronbach, L. J. (1970). *Essentials of Psychological Testing*, 3rd Edn. New York: Harper & Row.

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302.

Csikszentmihalyi, M., and Csikszentmihalyi, I. S. (Eds.) (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness.* New York, NY: Cambridge University Press.

Darnon, C., Dompnier, B., Gilliéron, O., and Butera, F. (2010). The interplay of mastery and performance goals in social comparison: a multiple-goal perspective. *J. Educ. Psychol.* 102, 212–222.

Datta, L. (1963). Test instructions and identification of creative scientific talent. *Psychol. Rep.* 13, 495–500.

Detterman, D. K. (1993). "The case for the prosecution: transfer as an epiphenomenon," in *Transfer on Trial: Intelligence, Cognition, and Instruction*, eds D. K. Detterman and R. J. Sternberg (Norwood, NJ: Ablex), 1–24.

Dickenson, T. L., Hassett, C. E., and Tannenbaum, S. I. (1986). *Work*

Performance Ratings: A Meta-Analysis of Multitrait-Multimethod Studies. San Antonio: Texas Maxima Corp.

Dumas, J. (1986). Controlling for autocorrelation in social interaction analysis. *Psychol. Bull.* 100, 125–127.

Ebbinghaus, H. (1913). *Memory: a Contribution to Experimental Psychology* (Henry A. Ruger and Clara E. Bussenius, Trans.). Originally published in New York by Teachers College, Columbia University. (Original German work "Über das Gedächtnis" published 1885). Available at: http://psychclassics.yorku.ca/Ebbinghaus/index.htm [Accessed on 13 January 2010].

Eckes, T. (1990). Situationskognition: Untersuchungen zur Struktur von Situationsbegriffen. *Z. Sozialpsychol.* 21, 171–188.

Edwards, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden.

Elms, A. C. (1975). The crisis of confidence in social psychology. *Am. Psychol.* 30, 967–976.

Endler, N. S. (1983). "Interactionism: a personality model, but not yet a theory," in *Personality – Current Theory and Research. Nebraska Symposium on Motivation*, 1982, ed. M. M. Page (Lincoln, NE: University of Nebraska Press), 155–200.

Endler, N. S. (1997). Stress, anxiety and coping: the multidimensional interaction model. *Can. Psychol./Psychologie canadienne* 38, 136–153.

Endler, N. S., Hunt, J. McV., and Rosenstein, A. J. (1962). An S-R inventory of anxiousness. *Psychol. Monogr.* 76 (17, Whole No. 536), 1–33.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *J. Pers. Soc. Psychol.* 37, 1097–1126.

Epstein, S. (1980). The stability of behavior. II. Implications for psychological research. *Am. Psychol.* 35, 790–806.

Feather, N. T. (1988). Values, valences, and course enrollment: testing the role of personal values within an expectancy-value framework. *J. Educ. Psychol.* 80, 381–391.

Fiske, D. W. (1982). "Convergent-discriminant validation in measurements and research strategies," in *Forms of Validity in Research*, eds Brinberg, D. and Kidder, L. H. (San Francisco: Jossey-Bass), 77–92.

Fiske, D. W., and Butler, J. M. (1963). The experimental conditions for measuring individual differences. *Educ. Psychol. Meas.* 23, 249–266.

Flanders, N. A. (1970). *Analysing Teaching Behavior*. Reading, MA: Addison-Wesley.

Follman, J. (1984). Cornucopia of correlations. *Am. Psychol.* 39, 701–702. (Comment)

Frederiksen, N. (1972). Toward a taxonomy of situations. *Am. Psychol.* 27, 114–123.

Freudenthaler, H., Spinath, B., and Neubauer, A. (2008). Predicting school achievement in boys and girls. *Eur. J. Pers.* 22, 231–245.

Geller, D. M. (1978). Involvement in role-playing simulations: a demonstration with studies on obedience. *J. Pers. Soc. Psychol.* 36, 219–235.

Gerhard, T. (2008). Bias: considerations for research practice. *Am. J. Health Syst. Pharm.* 65, 2159–2168.

Ginsburg, G. S., Grover, R. L., Cord, J. J., and Ialongo, N. (2006). Observational measures of parenting in anxious and nonanxious mothers: does type of task matter? *J. Clin. Child Adolesc. Psychol.* 35, 323–328.

Greeno, J. A., and the Middle School Mathematics Through Applications Project Group (1998). The situativity of knowing, learning, and research. *Am. Psychol.* 53, 5–26.

Griffo, R., and Colvin, C. R. (2010). Contextualized personality assessment: identifying situation dimensions associated with behavioral consistency and situation specific behavior. *Poster presented at the 11th Annual Meeting of the Society for Personality and Social Psychology. In SPSP: The 11th Annual Meeting of the Society for Personality and Social Psychology* (p. 434). Available at: http://www.spspmeeting.org/documents/SPSP2010_Program.pdf [Accessed on 23 January 2010].

Groeben, N., Wahl, D., Schlee, J., and Scheele, B. (1988). *Forschungsprogramm subjektive Theorien. Eine Einführung in die Psychologie des reflexiven Subjekts*. Tübingen: Francke.

Guba, E., and Lincoln, Y. S. (2000). "Epistemological and methodological bases of naturalistic inquiry," in *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, 2nd Edn, eds D. L. Stufflebeam, G. F. Madaus, and T. Kellaghan (Boston: Kluwer), 362–381.

Hartshorne, H., and May, M. A. (1928). *Studies in the Nature of Character*, Vol. 1, *Studies in Deceit*. New York: Macmillan.

Heiss, R. (1948). "Person als Prozess," in *Bericht über den Ersten Kongress des Berufsverbandes Deutscher Psychologen*. Bonn (1947). Hamburg: Noelke, 11–25. Reprinted in K. J. Groffmann and K.-H. Wewetzer (Eds.) (1968), Person als Prozess. Festschrift zum 65. Geburtstag von Robert Heiss. (Bern: Huber), 17–37.

Hensler, M., and Wood, D. (2010). Motives, abilities, and perceptions underlying variation in big five traits. *Poster presented at the 11th Annual Meeting of the Society for Personality and Social Psychology. In SPSP: The 11th Annual Meeting of the Society for Personality and Social Psychology* (p. 347). Available at: http://www.spspmeeting.org/documents/SPSP2010_Program.pdf [Accessed on 23 January 2010].

Heppner, P. P., Heppner, M. J., Lee, D. G., Wang, Y.-W., Park, H.-J., and Wang, L.-F. (2006). Development and validation of a collectivist coping styles inventory. *J. Couns. Psychol.* 53, 107–125.

Herber, H.-J., and Vásárhelyi, É. (2002). Lewins Feldtheorie als Hintergrundsparadigma moderner Motivations- und Willensforschung (im Vergleich zu Behaviorismus, Psychoanalyse, Gestalt- und Kognitionspsychologie). *Salzburger Beitr. Erziehungswissenschaft* 6, 37–100.

Herrmann, T. (1976). *Die Psychologie und ihre Forschungsprogramme*. Göttingen: Hogrefe.

Hertwig, R., and Ortmann, A. (2008). Deception in experiments: revisiting the arguments in its defense. *Ethics Behav.* 18, 59–92.

Hetherington, J. (1997). *Lecture 16: Advanced Research Design*. Available at: http://mccoy.lib.siu.edu/projects/psyc/hetherington/lect16.ppt [Accessed on 05 October 2005].

Higgins, R. L., Frisch, M. B., and Smith, D. (1983). A comparison of role-played and natural responses to identical circumstances. *Behav. Ther.* 14, 158–169.

Hoefert, H.-W. (Ed.) (1982a). *Person und Situation. Interaktionspsychologische Untersuchungen*. Göttingen: Hogrefe.

Hoefert, H.-W. (1982b). "Ansätze zu einer kompetenzspezifischen Situationstaxonomie," in *Person und Situation. Interaktionspsychologische Untersuchungen*, ed. H.-W. Hoefert (Göttingen: Hogrefe), 85–106.

Hofer, M. (1984). "Erziehungsleitende Zielvorstellungen von Lehrern," in *Jahrbuch für empirische Erziehungswissenschaft*, ed. G. Trommsdorff (Düsseldorf: Schwann), 105–126.

Hunt, J. McV. (1965). Traditional personality theory in the light of recent evidence. *Am. Sci.* 53, 80–96.

Jacob, T., Tennenbaum, D., Seilhamer, R. A., Bargiel, K., and Sharon, T. (1994). Reactivity effects during naturalistic observation of distressed and nondistressed families. *J. Fam. Psychol.* 8, 354–363.

Jones, E. E., and Nisbett, R. E. (1971). "The actor and the observer: divergent perceptions of the causes of behavior," in *Attribution: Perceiving the Causes of Behaviour*, eds E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, and D. Winer (Morristown, NJ: General Learning Press), 79–94.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., and Van Lange, P. A. M. (2003). *An Atlas of Interpersonal Situations*. New York: Cambridge University Press.

Kenrick, D. T., and Funder, D. C. (1988). Profiting from controversy: lessons from the person-situation debate. *Am. Psychol.* 43, 23–34.

Klehe, U.-C., Anderson, N., and Hoefnagels, E. A. (2007). Social facilitation and inhibition during maximum versus typical performance situations. *Hum. Perform.* 20, 223–239.

Klehe, U.-C., Anderson, N., and Viswesvaran, C. (2007). More than peaks and valleys: introduction to the special issue on typical and maximum performance. *Hum. Perform.* 20, 173–178.

Krahé, B. (1992). *Personality and Social Psychology. Towards a Synthesis*. London: Sage.

Krampen, G. (1984). "Methodologische Aspekte der Erfassung erziehungsleitender Vorstellungen," in *Jahrbuch für Empirische Erziehungswissenschaft*, ed. G. Trommsdorff (Düsseldorf: Schwann), 63–83.

Lanyon, R. I. (1984). Personality assessment. *Annu. Rev. Psychol.* 35, 667–701.

Lave, J. (1988). *Cognition in Practice*. New York: Cambridge University Press.

Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper & Row.

Magnusson, D., and Endler, N. S. (1977). "Interactional psychology: present status and future perspectives," in *Personality at the Crossroads: Current Issues in Interactional Psychology*, eds D. Magnusson and N. S. Endler (Hillsdale, NJ: Erlbaum), 3–31.

Mangos, P. M., Steele-Johnson, D., LaHuis, D., and White, E. D. III (2007). A multiple-task measurement framework for assessing maximum-typical performance. *Hum. Perform.* 20, 241–258.

Marcus, B., Goffin, R. D., Johnston, N. G., and Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Hum. Perform.* 20, 275–285.

Marsh, H. W. (1990). Confirmatory factor analysis of multitrait-multimethod data: the construct validation of multidimensional self-concept responses. *J. Pers.* 58, 661–692.

McCleary, R., and Hay, R. A. Jr., Meidinger, E. E., and McDowall, D. (1980). *Applied Time Series Analysis for the Social Sciences*. Beverly Hills, CA: Sage.

McGuire, W. J. (1969). "Suspiciousness of experimenter's intent," in *Artifact in Behavioral Research*, eds R. Rosenthal and R. L. Rosnow (New York: Academic Press), 13–57.

Mehrabian, A. (1971). Verbal and nonverbal interaction of strangers in a waiting situation. *J. Exp. Res. Person.* 5, 127–138.

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 371–378. Reprinted in Miller, A. G. (ed.) (1972). *The Social Psychology of Psychological Research* (New York: The Free Press/London: Collier-Macmillan), 82–105.

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychol. Rev.* 80, 252–283.

Mischel, W. (2004). Toward an integrative science of the person. *Annu. Rev. Psychol.* 55, 1–22.

Mischel, W. (2007). "Toward a science of the individual: past, present, future?" in *Persons in Context. Building a Science of the Individual*, eds Y. Shoda, D. Cervone, and G. Downey (New York: Guilford), 263–277.

Mischel, W., and Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychol. Rev.* 89, 730–755.

Mischel, W., and Peake, P. K. (1983). Some facets of consistency: replies to Epstein, Funder, and Bem. *Psychol. Rev.* 90, 394–402.

Mischel, W., and Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychol. Rev.* 102, 246–268.

Moser, K. (1991). *Konsistenz der Person*. Göttingen: Hogrefe.

Newcomb, T. M. (1929). *Consistency of Certain Extrovert-Introvert Behavior Patterns in 51 Problem Boys*. New York: Columbia University, Teachers College, Bureau of Publications.

O'Hara, L. A., and Sternberg, R. J. (2001). It doesn't hurt to ask: effects of instructions to be creative, practical, or analytical on essay-writing performance and their interaction with students' thinking styles. *Creat. Res. J.* 13, 197–210.

Ones, D. S., and Viswesvaran, C. (2007). A research note on the incremental validity of job knowledge and integrity tests for predicting maximal performance. *Hum. Perform.* 3, 293–303.

Orne, M. T. (1969). "Demand characteristics and the concept of quasi-controls," In *Artifact in Behavioral Research*, eds R. Rosenthal and R. L. Rosnow (New York: Academic Press), 143–179.

Patry, J.-L. (1982). "Laborforschung – Feldforschung," in *Feldforschung. Methoden und Probleme der Sozialwissenschaften unter natürlichen Bedingungen*, ed. J.-L. Patry (Bern: Huber), 17–42.

Patry, J.-L. (1989a). Contradictory goals, different expectations: towards an explanation of cross-situational specificity in social behavior. *Psychol. Rep.* 65, 1331–1339.

Patry, J.-L. (1989b). Evaluationsmethodologie zu Forschungszwecken – Ein Beispiel von "kritischem Multiplizismus". *Unterrichtswissenschaft* 17, 359–374.

Patry, J.-L. (1991a). *Transsituationale Konsistenz des Verhaltens und Handelns in der Erziehung*. Bern: Lang.

Patry, J.-L. (1991b). Der Geltungsbereich sozialwissenschaftlicher Aussagen: das Problem der Situationsspezifität. *Zeitschrift für Sozialpsychologie* 22, 223–244.

Patry, J.-L. (1992). A framework for the explanation of cross-situational inconsistencies in social behavior. *New Ideas Psychol. Int. J. Innov. Theory Psychol.* 10, 47–62.

Patry, J.-L. (1997a). Eine Person – mehrere Werte. Überlegungen zum intrapersonalen Wertpluralismus. *Pädagog. Rundsch.* 51, 63–81.

Patry, J.-L. (1997b). The lesson interruption method in assessing situation-specific behavior in classrooms. *Psychol. Rep.* 81, 272–274.

Patry, J.-L. (2000). "Kaktus und Salat – Zur Situationsspezifität in der Erziehung," in *Situationsspezifität in pädagogischen Handlungsfeldern*, eds J.-L. Patry and F. Riffert (Innsbruck: StudienVerlag), 13–52.

Patry, J.-L. (2004). Situation specificity, validity of the assessment, and the lab-field-problem. *Salzburger Beitr. Erziehungswissenschaft* 8, 1, 37–52. Available at: http://www.sbg.ac.at/erz/salzburger_beitraege/fruehling_2004/patry_1_04.pdf [Accessed on 5 February 2010].

Patry, J.-L. (2005a). "Intrapersonaler Wertepluralismus in der Erziehung: Theorie und konkrete Beispiele," in *Wertkonflikt und Wertewandel. Eine pluridisziplinäre Begegnung*, eds C. Giordano and J.-L. Patry (Münster: Lit), 135–150.

Patry, J.-L. (2005b). *Wissenschaftliche Schwerpunktsetzung*. Salzburg: Fachbereich Erziehungswissenschaft. Available at: http://www.uni-salzburg.at/pls/portal/docs/1/80064.PDF [Accessed on 16 February 2010].

Patry, J.-L. (2007). Lehrerinnen und Lehrer handeln situationsspezifisch – oder sie sollten. Eine Antwort auf auf Thonhausers Frage "Tun sie das?" *Salzburger Beitr. Erziehungswissenschaft* 11, 61–69. Available at: http://www.uni-salzburg.at/pls/portal/docs/1/553921.PDF [Accessed on 01 February 2010].

Patry, J.-L. (2008). "Konkurrenz, Koexistenz, Komplementarität qualitativer und quantitativer Methoden in der Erziehungswissenschaft aus der Perspektive des Kritischen Multiplizismus," in *Qualitative und quantitative Aspekte. Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung*, eds F. Hofmann, C. Schreiner, and J. Thonhauser (Münster: Waxmann), 133–150.

Patry, J.-L. (2009a). "Des Kaisers immer neue Kleider. Über das Finden und Vergessen von Theorien zur Situationsspezifität in der psychologischen Forschung," in *Trugschlüsse und Umdeutungen. Multidisziplinäre Betrachtungen unbehaglicher Praktiken*, eds C. Giordano, J.-L. Patry, and F. Rüegg (Berlin: Lit), 91–110.

Patry, J.-L. (2009b). "Nicht zu viel und nicht zu wenig: Grundlagen praktischen Tuns," in *Schule 2020 aus Expertensicht. Zur Zukunft von Schule, Unterricht und Lehrerbildung*, eds D. Bosse and P. Posch (Wiesbaden: Verlag für Sozialwissenschaften), 285–291.

Patry, J.-L., and Schrattbauer, B. (2000). Rollenkonflikte in der Bewährungshilfe. *Neue Prax* 30, 176–187.

Patry, J.-L., Schwetz, H., and Gastager, A. (2000). Wissen und Handeln. Lehrerinnen und Lehrer verändern ihren Mathematikunterricht. *Bildung Erzieh.* 53, 271–286.

Pauls, C. A., and Crost, N. W. (2005). Cognitive ability and self-reported efficacy of self-presentation predict faking on personality measures. *J. Individ. Differ.* 26, 194–206.

Peake, P. K. (1982). *Searching for Consistency: the Carlton Student Behavior Study*. Doctoral Disseration, Stanford University. Dissertation Abstracts, 43, Section 8, Part B, 2746.

Pervin, L. A. (1978). Definitions, measurements, and classifications of stimuli, situations, and environments. *Hum. Ecol.* 6, 71–105.

Peterson, D. R. (1968). *The Clinical Study of Social Behavior*. New York: Appleton-Century-Crofts.

Phillips, D. C. (1997). How, why, what, when, and where: perspectives on constructivism in psychology and education. Issues in education. *Contrib. Educ. Psychol.* 3, 151–194.

Phillips, D. C., and Burbules, N. C. (2000). *Postpositivism and Educational Research*. New York: Rowman & Littlefield.

Price, R. H. (1974). The taxonomic classification of behavior and situations and the problem of behavior-environment congruence. *Hum. Relat.* 27, 567–585.

Price, R. H., and Blashfield, R. K. (1975). Explorations in the taxonomy of behavior settings: analyses of dimensions and classifications of settings. *Am. J. Community Psychol.* 3, 335–351.

Price, R. H., and Bouffard, D. L. (1974). Behavioral appropriateness and situational constraints as dimensions of social behavior. *J. Pers. Soc. Psychol.* 30, 579–586.

Radford, M. (2006). Researching classrooms: complexity and chaos. *Br. Educ. Res. J.* 32, 177–190.

Renkl, A. (1996). Träges Wissen: wenn Erlerntes nicht genutzt wird. *Psychol. Rundsch.* 47, 78–92.

Reyna, C., and Weiner, B. (2001). Justice and utility in the classroom: an attributional analysis of the goals of teachers' punishment and intervention strategies. *J. Educ. Psychol.* 93, 309–319.

Rosenthal, R. (1976). *Experimenter Effects in Behavioral Research*. New York: Irvington.

Rotter, J. B. (1954). *Social Learning and Clinical Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Sackett, D. L. (1979). Bias in analytic research. *J. Chronic Dis.* 32, 51–63.

Sackett, P. R. (2007). Revising the origins of the typical-maximum performance distinction. *Hum. Perform.* 20, 179–185.

Sackett, P. R., Zedeck, S., and Fogli, L. (1988). Relations between measures of typical and maximum job performance. *J. Appl. Psychol.* 73, 482–486.

Salomon, G., and Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educ. Psychol.* 24, 113–142.

Schmitt, M. (1990). *Konsistenz als Persönlichkeitseigenschaft? Moderatorvariablen in der Persönlichkeits- und Einstellungsforschung*. Berlin: Springer.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Sherman, R., Nave, C., and Funder, D. (2010). Situational similarity and personality traits as predictors of behavioral consistency. *Poster presented at the 11th Annual Meeting* *of the Society for Personality and Social Psychology. In SPSP: The 11th Annual Meeting of the Society for Personality and Social Psychology* (p. 397). Available at: http://www.spspmeeting.org/documents/SPSP2010_Program.pdf [Accessed on 23 January 2010].

Shoda, Y. (2007). "From humunculus to a system: toward a science of the person," in *Persons in Context. Building a Science of the Individual*, eds Y. Shoda, D. Cervone, and G. Downey (New York: Guilford), 327–331.

Smith-Jentsch, K. A. (2007). The impact of making targeted dimensions transparent on relations with typical performance predictors. *Hum. Perform.* 3, 187–203.

Snyder, M., and Ickes, W. (1984). "Personality and social behavior," in *Handbook of Social Psychology*, Vol. II, *Special Fields and Applications*, eds G. Lindzey and E. Aronson (New York: Random House), 883–947.

Spector, P. (1989). Method variance as an artifact in self-reported affect and perceptions at work: myth or significant problem? *J. Appl. Psychol.* 72, 438–443.

Spinath, B., Spinath, F. M., Harlaar, N., and Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence* 34, 363–374.

SPSP (2010). *The 11th Annual Meeting of the Society for Personality and Social Psychology*. Available at: http://www.spspmeeting.org/documents/SPSP2010_Program.pdf [Accessed on 23 January 2010].

Sullivan-Marx, E. M. (2006). Directions for the development of nursing knowledge. *Policy Polit. Nurs. Pract.* 7, 164–168.

Sutton, R. E. (2004). Emotional regulation goals and strategies of teaching. *Soc. Psychol. Educ.* 7, 379–398.

Thonhauser, J. (2007). Lehrer/innen handeln situationsspezifisch – Tun sie das? *Salzburger Beitr. Erziehungswissenschaft* 11, 47–60. Available at: http://www.uni-salzburg.at/pls/portal/docs/1/553921.PDF [Accessed on 17 January 2010].

Vernon, P. E. (1964). *Personality Assessment. A Critical Survey*. London: Methuen.

Wallace, J. (1966). An abilities conception of personality: some implications for personality measurement. *Am. Psychologist* 21, 132–138.

Watson, R (1959). "Historical review of objective personality testing: the search for objectivity," in *Objective Approaches to Personality Assessment*, eds B. M. Bass and I. A. Berg (Princeton, NJ: Van Nostrand), 1–23.

Webb, E. J., Sechrest, L., and Campbell, D. T. (2000). *Unobtrusive Measures Revised*. Thousand Oaks, CA: Sage.

Weber, S. J., and Cook, T. D. (1972). Subject effects in laboratory research: an examination of subject roles, demand characteristics, and valid inference. *Psychol. Bull.* 77, 273–295.

West, S. G. (Ed.) (1983). Personality and prediction: nomothetic and idiographic approaches. Special issue. *J. Pers.* 51, 275–604.

Whitehead, A. N. (1967). *The Aims of Education and Other Essays*. New York: Macmillan (1929).

Wiggins, J. S. (1973). *Personality and Prediction: Principles of Personality Assessment*. Reading, MA: Addison-Wesley.

Willerman, L., Turner, R. G., and Peterson, M. (1976). A comparison of the predictive validity of typical and maximal personality measures. *J. Res. Pers.* 10, 482–492.

Williams, L. J., and Cote, J. A., and Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: reality or artifact? *J. Appl. Psychol.* 74, 462–468.

Witt, E., and Donnellan, M. B. (2010). Traits are more than just situational sensitivities. *Poster Presented at the 11th Annual Meeting of the Society for Personality and Social Psychology. In SPSP: The 11th Annual Meeting of the Society for Personality and Social Psychology* (p. 237). Available at: http://www.spspmeeting.org/documents/SPSP2010_Program.pdf [Accessed on 23 January 2010].

Woehr, D. J., and Arthur, W. Jr. (2003). The construct-related validity of assessment center ratings: a review and meta-analysis of the role of methodological factors. *J. Manag.* 29, 231–258.

Wolters, C. A., and Daugherty, S. C. (2007). Goal structures and teachers' sense of efficacy: their relation and association to teaching experience and academic level. *J. Educ. Psychol.* 99, 181–193.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 February 2010; Paper pending published: 09 February 2010; accepted: 23 January 2011; published online: 17 February 2011.

*Citation: Patry J-L (2011) Methodological consequences of situation specificity: biases in assessments. Front. Psychology **2**:18. doi: 10.3389/fpsyg.2011.00018*

This article was submitted to Frontiers in Quantitative Psychology and Measurement, a specialty of Frontiers in Psychology.