

Running head: SITUATIONAL JUDGMENT TEST OF PERSONAL INITIATIVE

**A Situational Judgment Test of Personal Initiative and its Relationship to
Performance**

Ronald Bledow and Michael Frese

Department of Psychology, University of Giessen, Germany

This manuscript is published in Personnel Psychology. It is a post-review pre-publication version of the manuscript. Please refer to Personnel Psychology for the article: Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. Personnel Psychology, 62, 229-258.

Acknowledgment. We would like to thank Christina Krause for conducting an open-ended survey about critical incidents of personal initiative. The Situational Judgment Test of Personal Initiative is based on this work. We would also like to thank all subject matter experts and employees who supported our efforts to construct and validate the Situational Judgment Test of Personal Initiative.

Correspondence and requests for reprints should be addressed to Ronald Bledow,
Department of Psychology, University of Giessen, Otto-Behaghel-Str. 10F, D-35394 Giessen,
Germany; ronald.bledow@psychol.uni-giessen.de

Abstract

Although situational judgment tests have been found to be valid predictors of performance, they have rarely been used to measure particular constructs. In this study we apply the situational judgment test method to the measurement of personal initiative, a construct defined as situated action. We used respondents' situated preferences in mental simulations of work scenarios as formative indicators of their overall level of personal initiative at work. Results from a validation study showed that the Situational Judgment Test of Personal Initiative (SJT-PI) had adequate validity and complemented a Likert-type self-report measure of personal initiative in predicting behavioral criteria. Situated preferences for personal initiative were hypothesized to be proximal predictors of actual behavior and were accordingly found to mediate the relationship between generalized self-efficacy, felt responsibility and actual behavior. Furthermore, situated preferences for personal initiative could be differentiated empirically from organizational citizenship behavior. We conclude that situational judgment tests are a promising method for measuring personal initiative and may be a general means of improving the validity of measurement in organizations.

A Situational Judgments Test of Personal Initiative and its Relationship to Performance

Since Motowidlo, Dunnette and Carter (1990) first introduced SJTs as low fidelity simulations a large body of research and practical experience has emerged, demonstrating the validity and practical usefulness of SJTs (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). However, past work on SJTs has been mostly atheoretical focusing on the prediction of external criteria. Many authors are calling for a shift of focus towards the processes and constructs underlying situational judgments for the better understanding and improved application of SJTs (e.g. Weekley & Ployhart, 2005). In particular, although there seems to be agreement that SJTs can be used to measure different constructs (Schmidt & Chan, 2006), most previous attempts to develop SJTs measuring particular constructs have had only limited success (e.g. Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). In this paper we argue that SJTs are particularly useful for the measurement of constructs that are defined as situated action. We present a novel theoretical rationale and a practical approach for developing construct specific SJTs. The focal construct to which we apply this approach is personal initiative at work. By providing a SJT of personal initiative we intend to overcome limitations inherent in past research measuring personal initiative only with traditional Likert-type scales. We examine the conceptual and empirical relationship of a SJT and a Likert-type scale of personal initiative and their relationship with behavioral criteria. Thus, we intend to contribute both to an improved measure of personal initiative and to an improved concept of construct-based SJTs.

Personal initiative is an active performance concept stressing that people self-start to bring about positive individual and organizational outcomes (Fay & Frese, 2001). This aspect of performance has been neglected in traditional approaches to work performance (Griffin, Neal, & Parker, 2007) and there is a misfit between the theoretical concept and how it is measured in

organizational research (Frese, Fay, Hilburger, & Leng, 1997). To be successful on today's global markets companies need employees who actively attack problems, search for new opportunities and continuously improve their work environment. Companies with employees who simply do what they are told are losing their competitive edge (Frese & Fay, 2001). Fay and Frese (2001) defined Personal Initiative as "...work behavior characterized by its self-starting nature, its proactive approach, and by being persistent in overcoming difficulties that arise in the pursuit of a goal" (p. 133). To select individuals with high initiative, develop employees' competencies to show initiative, and foster an environment in which initiative behavior is supported companies require valid measures of the phenomenon. Frese et al. (1997) recommended a situational approach, in their case a situational interview (Latham & Saari, 1984) and warned against measuring personal initiative only with traditional self-report scales. Nevertheless, research on personal initiative and related constructs has mostly relied on Likert-type self-report scales (e.g. Frese, Teng, & Wijnen, 1999; Morrison & Phelps, 1999; Van Dyne & LePine, 1998; Bateman & Crant, 1993).

A Theoretical Approach to a Situational Judgment Test of Personal Initiative

What are the limitations of Likert-type self-report scales for the measurement of personal initiative? McClelland (1987) and Spencer & Spencer (1993) proposed that self-report questionnaires measure self-concepts that do not necessarily reflect actual behavior. Frese et al. (1997) argued that traditional self-report measures of personal initiative capture the importance people assign to personal initiative but not the extent of actual initiative at work. Items of the personal initiative scale are generalized statements about how an individual approaches work (e.g. "I actively attack problems at work"). They are not linked to specific situations at work and do not provide behavioral examples. Thus they constitute inferential measurements. Respondents

have to infer their standing on the dimension of personal initiative from their previous behavior at work or directly refer to their self-concept. If respondents try to anchor their responses in past behavior, the behavioral examples will be different among respondents depending on their idiosyncratic experience. Furthermore, respondents may have different notions about what constituted high or low initiative and therefore relate to different anchor points on Likert-type rating scales. Similar concerns about the use of Likert scales have been raised in cross-cultural research (Peng, Nisbett, Wong, 1997; König, Steinmetz, Frese, Rauch, Wang, 2007).

These general limitations of Likert-type items are particularly problematic for the measurement of personal initiative because personal initiative is defined on the level of observable, situated action. Theoretical developments on situated action (Suchman, 1987) and action theory (Frese & Zapf, 1994; Hacker, 2003) emphasize that researchers should not abstract action away from its circumstances but should “study how people use their circumstances to achieve intelligent action.” (Suchman 1987, p 50). From the perspective of situated action human activity emerges “out of the particularities of a given situation” (Nardi, 1996, p. 36). Human action cannot be conceptualized merely as the manifestation of decontextualized trait characteristics (Costall & Leudar, 1996). Human action is shaped by the constraints and affordances (i.e. action possibilities) of a given situation that interact with characteristics of the individual. Individuals are thus characterized by their distinctive patterns of variability in their actions across different situations (Mischel, 2004). The notion that the situational context is a meaningful source of variation in individual behavior is also supported by recent research on assessment centers. Lance et al. (2000) demonstrated that exercise factors (i.e. situational factors) are a valid source of variance in applicants’ performance rather than a source of systematic error variance.

By regarding personal initiative as situated action we emphasize the interrelationship between personal initiative and its context of performance. This implies that personal initiative is a domain specific activity and situated in the work setting (Frese et al, 2001). As it is a resource intensive activity (Bolino & Turnley, 2005) people tend to not show consistent levels of initiative across different domains in life. For instance, personal initiative is only weakly related to initiative taking outside work like being active in associations (Frese, 1997). Theoretically, we expect individuals to show varying levels of initiative even within the work setting depending on the particularities of a given situation. For instance, situations vary in the competencies they demand for acting (Mischel & Shoda, 1995). Even if a person has the proclivity to taking initiative at work he or she may not be active in a situation because he or she lacks the specific competencies needed to show initiative or because the person has a history of negative responses to use personal initiative in a particular situation. We think the embeddedness of personal initiative in work situations holds important implications for the measurement of personal initiative. There is a misfit between the theoretical notion of personal initiative as situated action and its abstract way of being measured with Likert-type scales in current organizational research. We argue that measurement of personal initiative can be improved if respondents communicate their preferences for specific actions grounded in simulated situations.

The Situational Judgment Test of Personal Initiative (SJT-PI) presents descriptions of work related situations to respondents and asks them to mentally simulate that they are faced by these situations. For each situation respondents are instructed to select among different actions the ones they would most likely and least likely perform. The situations refer to critical incidents of personal initiative. The actions vary in the degree to which they are high or low in personal initiative. Actions high in personal initiative are self- starting, proactive and overcome barriers.

Features of actions low in personal initiative are: taking the conventional path, accepting the status quo, and managing one's emotions rather than changing the situation (Frese & Fay, 2001).

Figure 1 provides a sample item.

Insert Figure 1 about here

The SJT-PI is a more indirect and situated measure than the Likert-type scale of personal initiative. We argue that respondents' decisions in mental simulations of work situations reveal their situated behavioral preferences for high or low initiative actions. Whereas the Likert-type measure of personal initiative asks respondents directly to report their level of personal initiative, the SJT-PI does not require respondents to infer their standing on the dimension of personal initiative. When respondents answer to the SJT-PI, they imagine being in the presented situations and select among different actions the ones they are most and least likely to perform. By making these decisions respondents indirectly communicate their situated preferences for high or low initiative actions. A respondent's overall preference for personal initiative is a function of respondent's situated preferences for personal initiative in different scenarios.

Motowidlo, Hooper and Jackson (2006) recently introduced the concept of implicit trait policies as an explanation of how individual trait differences affect SJT responses. In many ways we agree with this description. In Motowidlo et al's (2006) terms, people with extreme standings on a given trait accentuate trait relevant differences in SJT response options and link them to behavioral effectiveness. For people with a strong preference for or against showing initiative at work, the different levels of personal initiative in the SJT response options should be particularly salient and guide their judgments. People with a strong preference for initiative link high

initiative actions with behavioral effectiveness and are likely to select these actions when asked what they would do. People with a very low preference for initiative should also accentuate differences in levels of personal initiative but see more passive means of dealing with the situation as appropriate. However, there are also points related to the situatedness of action where we disagree with Motowidlo et al's (2006) approach. We are skeptical about a pure trait explanation of SJT responses. Motowidlo et al (2006) used SJTs as a method for assessing implicit components of personality. While we concur that SJTs constitute more indirect ways of measuring individual differences than Likert-type items, we stress that their potential for improving measurement in organizational research lies in their situatedness. Respondents take the situational context into account when answering each item. Responses thus emerge from the interaction of the person and the simulated situations.

Development of Hypotheses

We assume situated behavioral preferences in the low-fidelity simulation of the SJT-PI to be related to behavioral preferences that guide decision making of how to act at work. When answering the SJT-PI situated behavioral preferences influence the selection among different actions. "In situ", situated behavioral preferences shape the development of goals and intentions (Latham & Skarlicki, 1995), their transformation into concrete ideas of how to act in a given situation (Gollwitzer, 1999) and ultimately actual behavior (Latham & Saari, 1984).

Supervisors observe this situated behavior at work. They observe some people to prefer not taking initiative action at all, some people to take initiative in a few situations and others to take initiative in many situations. Based on these observations they generalize about employees' level of initiative. Situated behavioral preferences for initiative as measured with the SJT-PI should therefore be related to personal initiative at work as observed by supervisors.

Hypothesis 1: The SJT-PI is positively related to supervisor ratings of personal initiative.

Personal initiative is conceptualized as an active performance concept (Frese & Fay, 2001). It is expected to bring about long-term positive changes for individuals and organizations. In support of this reasoning numerous studies have linked personal initiative to various individual and organizational level outcomes. For instance, unemployed persons with a high degree of personal initiative find a job faster than those with low personal initiative (Frese et al., 1997). Several studies found a relationship between personal initiative and entrepreneurial success in different economic environments (Zempel, 1999; Koop, De Reu, & Frese, 2000; Krauss, Frese, Friedrich, 2007). Van Dyne and LePine (1998) showed that voice – a construct related to personal initiative – was significantly related to estimates of individual performance by peers, by supervisors and by self-assessment. We thus hypothesize situated behavioral preferences for personal initiative as precursors of action to be positively related to supervisors' overall performance ratings.

Hypothesis 2: The SJT-PI is positively related to supervisor ratings of overall performance.

We expect that the SJT-PI complements Likert-type self-report scales because it indirectly assesses situated behavioural preferences that are not part of the general self-concept. Both measures tap different information about respondents. While the situational judgment test assesses situated preferences, the Likert-type scale assesses a generalized self-concept related to personality. Past research has found the Likert-type scales of personal initiative to be highly related to proactive personality (uncorrected $r = .76$, Fay & Frese, 2001; Crant, 2000). Although we have argued that a situational measure of personal initiative is more in line with the theoretical concept, the more abstract and direct Likert-type scale of personal initiative has an

advantage. It is much broader in scope and can capture cognitions and facets of behavior that do not fall into the range of situations included in the SJT-PI. Therefore we expect both measures to complement each other and predict independent variance in behavior. Concerning the relationship between the Likert-type scale and the SJT-PI we expect only a moderate positive relationship. Even though both measures were intended to measure personal initiative, we think that different cognitions are involved in answering both measures and respondents reveal different facets of their level of personal initiative. This reasoning is similar as the reasoning on direct and more indirect measures of constructs, for instance achievement motivation (Brunstein, 2004). Both types of measures of a construct tend to correlate only moderately and are complementary in the prediction of behavioral criteria (Bing, 2006).

Hypothesis 3a: The SJT-PI is positively - but only moderately - related to self-ratings of personal initiative.

Hypothesis 3b: The SJT-PI and self-ratings of personal initiative independently predict personal initiative at work and overall performance.

Even if both measures of personal initiative are only moderately related they should still have similar relationships with constructs that are conceptually and empirically closely linked to personal initiative. Morrison & Phelps (1999) discussed felt responsibility and Speier & Frese (1997) generalized efficacy beliefs as leading to personal initiative. Building on this model we hypothesize generalized self-efficacy and felt responsibility to be related to both situated preferences for personal initiative and the self-concept of personal initiative. Furthermore we expect both operationalizations of personal initiative to mediate the relationship between self-efficacy and felt responsibility and personal initiative as observed by the supervisor. The path model is depicted in Figure 2.

Insert Figure 2 about here

Generalized self-efficacy refers to an employee's belief about his or her capacity to perform. High self-efficacy leads to higher goals and to greater persistence if difficulties arise (Bandura, 1991). Employees with high self-efficacy beliefs will tend to attach a higher likelihood of success to personal initiative. Three independent groups of researchers found a positive relationship between generalized self-efficacy and proactive, change oriented behavior (Speier & Frese, 1996; Morrison, & Phelps, 1999; Parker, Williams, & Turner, 2006). Felt responsibility is a further construct that has been shown to be related to personal initiative. Felt responsibility is "an individual's belief that he or she is personally obligated to bring about constructive change" (Morrison, & Phelps, 1999, p. 407). To the extent that people feel highly responsible for change, they will prefer high initiative actions and eventually take initiative because it is in line with their subjective norm and they attach positive valence to the outcome of taking initiative.

Hypothesis 4a: Generalized self-efficacy and felt responsibility are positively related to the SJT-PI and self-ratings of personal initiative.

We further hypothesize self-ratings of personal initiative and situated preferences for personal initiative to be mediators in the model. Only if psychological orientations like generalized self-efficacy and felt responsibility result in action can initiative behavior be observed by others. This active tendency to self-start and act is reflected in the self-ratings of personal initiative and situated preferences for personal initiative.

Hypothesis 4b: The SJT-PI and self-rated personal initiative mediate the relationship between the antecedents felt responsibility and generalized self-efficacy and personal initiative as observed by the supervisor.

In summary, the model (see Figure 2) predicts the same mediational position in a nomological network of related constructs for both the SJT-PI and self-ratings of personal initiative. We test this model to examine if the newly developed SJT-PI captures the performance domain it is intended to measure (Schmitt & Chan, 2006). To ensure discriminant validity – i.e. the SJT-PI should not measure constructs other than personal initiative - we examine its relationship with organizational citizenship behavior (OCB).

OCB is a concept that shares some features with personal initiative (Organ, 1997). Both organizational citizenship behavior (OCB) and personal initiative go beyond employees' role requirements and both are considered to contribute to organizational effectiveness. Two dimensions underlying OCB have been studied most frequently: altruism and compliance (Smith, Organ, & Near, 1983). Altruism is manifested in helping behavior, compliance in employees' conscientiousness. Helping behavior overlaps with personal initiative only if it is self-started and if there is a long-term focus (Frese & Fay, 2001). Showing someone an effective work procedure so that he or she can cope with future demands would be an example of both personal initiative and altruism. However, personal initiative is not necessarily altruistic. If employees pursue self-serving goals, this can be still considered personal initiative if it does not damage the organization (Frese & Fay, 2001). Compliance – the second important facet of OCB – is not related to personal initiative because it has a somewhat passive connotation and focuses on adherence to rules; in contrast, personal initiative often implies ignoring rules or even being somewhat rebellious toward existing rules (Frese & Fay, 2001). Although both personal initiative

and conscientiousness are conceptually distinct, both can be considered expressions of a need for achievement and may therefore be empirically related (Frese et al., 1997). Although there are theoretical reasons to expect a certain degree of overlap between the OCB dimensions of helping and conscientiousness and personal initiative, the concepts are theoretically distinct. With respect to the construct validity of the SJT-PI, it should be sufficiently distinguishable from organizational citizenship behavior. We predict the SJT-PI to be more strongly related to other measures of personal initiative than to measures of helping and conscientiousness.

Hypothesis 5a: The SJT-PI is more strongly related to self- and other reported personal initiative than to self- and other reported helping behavior.

Hypothesis 5b: The SJT-PI is more strongly related to self- and other reported personal initiative than to self- and other reported conscientiousness.

Method

Development of the Situational Judgment Test of Personal Initiative

We followed the general guidelines outlined by MacKenzie, Podsakoff & Jarvis (2005) for developing, modeling, and validating the SJT-PI as a measure of a construct with formative (or causal) indicators. We complemented the standard developmental procedure for SJTs (Motowidlo et al., 1990) with a theory driven approach. A SJT developed to assess a particular behavioral construct poses different requirements than SJTs that are solely constructed to predict an external criterion (Ployhart & Ryan, 2000). Response options of one item need to represent different levels of the same construct not entirely different constructs. Furthermore, since SJT-PI items are formative indicators of respondents' level of personal initiative, the range of situations

included is critical. Only if a heterogeneous and representative sample of work situations is included does the test measure personal initiative comprehensively.

As the first step of test development, we used an open-ended survey to collect critical instances of personal initiative that had actually occurred. In a second step, we constructed items based on the collected examples and the theoretical concept of personal initiative. In a third step, we let subject matter experts rate the degree of personal initiative for each response option and examined interrater agreement to ensure content validity of the SJT-PI.

The goal of the first step was to create a representative and heterogeneous pool of examples of personal initiative in the workplace. We choose an empirical approach and collected a sample of situations and behaviors that had actually occurred. We sent out an open-ended survey to participants ($N = 30$) working fulltime in various fields (e.g. teachers, salespersons, bank accountants). Participants were asked to think of a situation where a coworker or they themselves had tried to implement an improvement in their workplace, to describe the initial situation and to describe what had been done to improve it. Participants were also instructed to mention whether barriers had arisen while solving the problem and what had been done to overcome those barriers. In addition, participants indicated whether the behavior was formally required. This ensured that only self-started behaviors not prescribed by formal job requirements were included in the test.

The goal of the second step was to construct preliminary SJT-PI items. A total of 25 situations could be derived as stems for the item pool. The situations' demand to show initiative was generally high which might have led all respondents to indicate high personal initiative. For the purpose of differentiating between high and low initiative respondents, we modified the situations based on the competency demand hypothesis. The competency demand hypothesis

(Mischel & Shoda, 1995) states that situations vary in the degree to which individual differences in a disposition will lead to individual differences in behavior. Situations with high competency demands relevant for a given trait will differentiate between people high or low on this disposition. Situations with low competency demands will not. Since the differentiation in situated preferences for personal initiative was the objective of the SJT-PI, the empirically derived situations were modified to place higher demands on showing initiative. A defining feature of personal initiative is overcoming barriers and persisting in the face of difficulties. We thus included a barrier in each description of the situation that places high demands on showing initiative and makes “taking the usual path” (i.e. low initiative) a reasonable and acceptable alternative. Examples of barriers were resistance of others, cost, and time pressure. The situations resembled the dilemmas presented in the situational interview (Latham & Skarlicki, 1995). A dilemma suggests two different ways of action. Either way has some disadvantage. We thought that this framing makes the “best” way of dealing with a situation less obvious and, thereby, reduces the problem of socially desirable answers (Peeters & Lievens, 2005).

Next, a first response option high in personal initiative was formulated for each item on the basis of the examples collected in step 1. In order to construct response options low in initiative with high social desirability, the response options were framed as high in emotional stability (e.g. not getting upset, staying calm and patient). Whereas response options high in personal initiative were aimed at actively changing the situation that was presented, response options low in personal initiative were aimed at coping with the situation by adapting to it (e.g. regulating one’s emotions). Thereafter, response options were constructed which fall in the middle range between the high and the low personal initiative response options. They were derived from the high and the low personal initiative response options to ensure that the stated

behavior differed mainly on the dimension of personal initiative. An example of a behavior that falls in the middle range was making others aware of a problem but not solving it. All SJT-PI items had four or five response options that were categorized as either high (+1), medium (0) or low (-1) in personal initiative. We arrived at a list of 13 items that were sufficiently general to apply them in a wide variety of jobs and reflected the heterogeneity of examples collected in the open-ended survey.

The goal of the third step was to ensure content validity of SJT-PI items by examining interrater agreement of the level of personal initiative in the response options of each item. In keeping with the construct approach, we asked subject matter experts who were familiar with the psychological concept of personal initiative to rate the different levels of personal initiative. We selected a group of 25 graduate students of industrial and organizational psychology familiar with the construct of personal initiative as subject matter experts. They had not been involved in the process of item development. We asked them to indicate for each response option of the 13 items if it was low, medium or high in personal initiative. We used the average deviation (AD) index to examine agreement among raters (Burke, Finkelstein, & Dusig, 1999). The AD-index quantifies the average dispersion of ratings around the mean of all ratings. Lower AD-values indicate higher agreement among raters. Burke and Dunlap (2002) have provided a criterion for acceptable levels of agreement. For a three-point scale the upper-limit for acceptable agreement of ratings is .50. The response options for all but one SJT-PI item were below the value of .50. We discarded the item with a lack of agreement. The averaged AD values for the remaining 12 items ranged between .12 and .44.

The scoring procedure for respondents' answers was based on the procedure as recommended by Motowidlo et al. (1990). Respondents were asked to indicate which of the 4 or

5 response options presented for each item they would perform most likely and least likely. For the most likely ratings respondents received +1 if they selected the response options high in personal initiative, 0 if they selected a response option with a medium level of personal initiative and -1 if they selected a response option low in personal initiative. Responses for the question asking about what respondents' would do least likely were scored inversely: if a response option low in personal initiative was selected as what would be done least likely, the score was +1; for the response option with medium levels of personal initiative the score was 0; for the response option with high levels of personal initiative the score was -1. The most and least likely ratings were combined for each item resulting in a score that could vary on a 5 point scale between -2 (low initiative) and +2 (high initiative) (McDaniel & Nguyen, 2001). Respondents' overall scale score on the SJT-PI was derived from the mean of item scores.

Situated Preferences as Formative Indicators of Personal Initiative

From our theory of SJTs as consisting of items that are simulations of situated action, follow a number of methodological implications. The most important one is that item responses are due to interactions of persons and simulated situations. This implies that situational aspects of each scenario influence responses and therefore SJT-PI items are not parallel measures of a unidimensional construct. Each scenario item of the SJT-PI measures a unique phenomenon: The preference for high or low initiative within a specific context. An item response emerges out of the interaction between a person and a simulated situation. For instance, respondents who report to take initiative in many situations may not choose response options high in personal initiative in some situations: they may lack the specific skills they need to take initiative or they may have a history of negative responses in their work to use personal initiative in a particular situation. Variability in responses across different scenarios is therefore expected within our situated

approach and has methodological implications for the relationship between scenario items and the overall score on the SJT-PI.

As we do not assume a unidimensional latent variable of personal initiative as the single data generating mechanism across all scenarios, a reflective indicator model appears not to be in line with our theoretical approach. SJT-PI items are formative indicators that define the overall level of personal initiative. Conceptually each item has its own latent variable which represents the true behavioral preference of a person in a specific simulated situation (MacKenzie et al., 2005). The overall strength of employees' preference for personal initiative in this low-fidelity simulation is determined by the situated preferences in all situations. The same logic holds for how the actual extent of personal initiative is defined "in situ", as a function of relevant behaviors employees perform across all relevant situations at work (Frese & Fay, 2001). Therefore, the overall score on the SJT-PI is a composite measure of how frequently respondents report situated preferences for high initiative behaviors. The overall score does not represent a unidimensional latent trait which is conceived as residing within individuals influencing their judgments.

Defining situated preferences as formative indicators leaves open the possibility that there are situational factors that influence multiple items. Formative indicator items can share the same antecedents but also have antecedents that affect only a single item or a group of items (Bollen & Lennox, 1991)¹. Theoretically, we expect that behavioral preferences are more consistent if the situational context of personal initiative is similar (Mischel & Shoda, 1995). To explore this issue we grouped scenario items into three content domains and applied factor analysis. The three domains are personal initiative directed at improving organizational functioning, personal initiative directed towards improving one's own working conditions and personal initiative that

required overcoming the resistance of supervisors and colleagues. Factor analysis of the SJT-PI items extracted three factors that correspond to the three content domains with four items loading on each factor. In line with the theoretical notion of situational specificity, common factors explained only a small proportion of item variance and factor loadings were small ranging between .24 and .63. This indicates that item responses are influenced by factors common to groups of items as well as unique factors. Selecting only items with high loadings on common factors would not be in line with our theoretical approach because factors unique to one item can carry meaningful information about the overall level of personal initiative.

To test our approach we modeled SJT-PI items as formative and self- and supervisor ratings of personal initiative and performance as reflective indicators of personal initiative. Thus, the latent personal initiative construct is not by itself reflective or formative. The terms describe the relationships between measures and a construct, not the nature of the construct (Edwards & Bagozzi, 2000). Before testing the model, we formed three parcels of SJT-PI items. We used parceling for the methodological reasons of reducing the number of indicators, reducing nonnormality of indicators and avoiding multicollinearity among indicators. Items were combined to parcels that belonged to same content domain and loaded on the same factor. However, parcels should not be interpreted as situational factors. We also report results for a model without parceling of items.

To test structural relationships of the SJT-PI with other variables we used the composite of the twelve items as a manifest variable. By using the average score, we determined that each items contributes with the same weight to the overall score. The composite construct is thus not dependent on criterion constructs. In contrast, in the latent model with formative indicators, the scenarios are weighted so that the resulting latent construct best predicts the criteria (self-ratings

of personal initiative, supervisor ratings of personal initiative and performance) (Edwards & Bagozzi, 2000; Howell, Breivik, & Wilcox, 2007). By using the average score with equal weights we do not empirically maximize predictive validity. As the 12 SJT items of the composite are not unidimensional and influenced by general as well as context specific factors internal consistency is not expected and Cronbach's alpha is not the appropriate measure of reliability (Edwards & Bagozzi, 2000; MacKenzie et al., 2005). Instead, we used test-retest reliability to examine the reliability of the SJT-PI (Motowidlo et al. 1990).

Qualitative Pilot Study on Item Response Strategies

To explore the cognitive processes involved in responding to the different measures of personal initiative we used the thinking aloud method (Ericsson & Simon, 1980). Six subjects working in different jobs were asked to answer four items of the Likert-type scale of personal initiative and four items of the SJT-PI. Respondents were repeatedly asked to think aloud while answering to the items. After answering to the items respondents were asked to explicitly state how they generated their responses. All interviews were recorded and respondents' comments were transcribed verbatim.

Content analyses of respondents' comments were done. When asked to indicate their level of initiative on Likert-type items (e.g. *I take initiative immediately even if others don't; I use opportunities quickly in order to attain my goals*) respondents used two different sources of information: examples of past behavior and generalized thoughts about themselves. Some respondents used only one source of information others used both sources. Examples of thoughts respondents expressed while answering the self-report items of personal initiative are: "I cannot work with problems on my mind that's why I need to get rid of them"; "in general I am an active person"; "yes, certainly if there is an opportunity, I take it"; "at work I am definitely the one who

thinks most about how to solve problems”. Some respondents provided specific examples of past initiative others had difficulties in grounding their responses in specific examples or did not use any examples at all (“I take initiative sometimes but I cannot remember an example at the moment”). Respondents also commented on some difficulties they had when answering the items: “it depends on the situation”; “it really depends on what kind of task I am dealing with”; “the question is what does taking initiative mean, I can’t deal with this question”. Some respondents thus perceived items as ambiguous and very general.

When answering SJT-PI items respondents expressed different kinds of thoughts. They did not refer to past behavior or how they perceived themselves. Rather they imagined to be in a similar situation at their current job and reflected about how they would act. (“I tried to visualize the situation and thought how I would act”). In most cases some response options were directly excluded from further consideration because respondents perceived them as inappropriate or something they would certainly not do (“this does not make sense”; “I don’t think this is the right thing to do”; “that is not enough, I need more control”). After having excluded some response options, respondents compared the remaining options (“It is difficult for me to decide between those two”). They frequently based their decision on the consequences they thought an action would have and selected the response option they thought would bring about the outcomes they valued (“this would just create more conflict”). In some cases respondents thought about what they would do in a situation before reading the response options and then selected the option that was closest to what they had generated. All thoughts respondents expressed about how to act were closely linked to the situation presented by the scenarios. They did not try to generalize away from the situation but mentally observed themselves in the simulated situations. Thus, this small qualitative pilot study was in line with our theory that answers to SJT-items are based on

mental simulations of situated actions, while answers to Likert-type items are often based on the general self-concept or available examples. Likert-type items of personal initiative might thus be biased by ego-defenses and availability bias (Tversky & Kahneman, 1973).

Sample of the Validation Study

The validation study was conducted with employees and supervisors of six regional banks in Germany. In total, 140 employees were asked to participate in the study. 126 employees filled out the survey (response rate of 89%). The average age of respondents was 36 years, the average organizational tenure was 14 years and the average job tenure was six years. 31 % of participants were women. In terms of hierarchical level, 57 % indicated that they held non-managerial positions, 18% held lower-level management positions, 23% were in middle management and 2% worked in upper management. We obtained ratings from 22 supervisors for 77 employees. Two banks were not willing to provide supervisor ratings because of internal policies regarding performance assessments of individuals. On average, the supervisors were responsible for 25 employees and were 44 years old. 19 supervisors were male, three supervisors were female. Supervisors had been responsible for at least six months for the employees for whom they provided ratings. We compared the samples of employees with and without supervisor ratings with multiple t-tests. There were no significant differences for demographic variables, work related variables (e.g. position, tenure) and the key constructs of our study. Generalization of results from analyses with the sub-sample of employees with supervisor ratings to the entire sample seems justified.

Survey Instruments

Data were collected from two sources: self-report and supervisor ratings. Employee ratings were collected with an online survey. Supervisor ratings were collected with a paper and pencil questionnaire.

Employee Self-Report Ratings. *Situational Judgment Test of Personal Initiative (SJT-PI).* The construction of the scale was described above. In all respondents' overall score on the SJT-PI was derived from the mean of the 12 item scores. As described before, our theory of situated item responses and the formative indicator approach are incompatible with using Cronbach's Alpha as an reliability estimate for the overall score of the SJT-PI (Motowidlo et al., 1990; MacKenzie et al., 2005).² We, therefore, estimated test-retest reliability with a subset of four items and 39 subjects. The time lag between the first and the second wave of data collection was approximately 11 weeks. The test-retest reliability (i.e. Pearson correlation) of the aggregated four items was .73 and can be regarded as acceptable (Nunnally, 1978).

Self-Reported Personal Initiative. The seven Likert-type item scale by Frese et al. (1997) was used to measure employees' self-reported level of personal initiative in the work place. The answer scale's internal consistency was $\alpha = .80$, an example item is "I actively attack problems".

Helping. The organizational citizenship behavior scale *helping* measures employees' helping behavior and is an operationalization of altruism. The scale was developed by Organ and Konovsky (1989) and adapted by Van Dyne and LePine (1998). We used five items of the scale. Cronbach's Alpha was .80. An example item is: "I am always willing to help and support others".

Conscientiousness. We used the four-item scale *Conscientiousness* (Farh, Podsakoff, & Organ, 1990) to measure a second facet of organizational citizenship behavior. Since Cronbach's alpha for the four item scale was insufficient, we used only a single marker item that best

represented the construct. We used the item that asked employees to directly assess their level of conscientiousness (“I am one of the most conscientious employees of my workgroup”). Because a single item does not sufficiently capture the performance domain of conscientious behavior, results should be interpreted cautiously.

Generalized Self-Efficacy. We used the generalized self-efficacy scale developed by Sherer (1982). Instead of the original scale with 12 items, we used a shorter scale with 6 items that were sufficiently reliable in a previous study with a highly similar sample (Schmitt, 2004). The scale’s reliability was $\alpha = .79$. An example item is “I feel insecure about my ability to do things.”

Felt Responsibility. The *felt responsibility* scale was developed by Morrison und Phelps (1999) to assess the responsibility employees perceive for improving their work environment. The scale consisted of five items with internal consistency of $\alpha = .77$. An example item is “I feel a personal sense of responsibility to bring about change at work”.

Supervisor Ratings. Overall Performance. Supervisor ratings of *overall performance* were measured with three items developed by Motowidlo and Van Scotter (1994). For each item supervisors made their ratings on a seven-point scale. Behavioral anchors for the lower (1,2), middle (3-5) and upper (6,7) range of the scales were used. A sample item asked whether the employee contributes less, an average amount or more to the performance of the department than most other members. Cronbach’s alpha of the aggregated overall performance measure was .96.

Personal Initiative. The scale for supervisors’ ratings of personal initiative consisted of seven items. Frese et al. (1997) demonstrated the scales’ convergent validity with ratings of personal initiative by others and other operationalizations of personal initiative. In this study the

scales' internal consistency was $\alpha = .91$. A sample item is: "This employee actively attacks problems".

Helping and Conscientiousness. The same items of the *helping* and *conscientiousness* scales employees had answered were also rated by supervisors, the only difference being the referent ("this employee" rather than "I"). Internal consistencies of the *helping* and *conscientiousness* scales were $\alpha = .89$, $\alpha = .81$, respectively

Control Variables. Respondents provided information about their job position, age and gender.

Results

The three item parcels of the model presented in Figure 3 define the overall level of personal initiative as measured with the SJT-PI. This emergent variable is a function of respondents' situated preferences which is represented in the model by the arrows pointing in direction of the variable. In contrast to traditional factor models, this emergent variable is thus not a latent variable within individuals that influence item responses (Borsboom, Mellenbergh, & van Heerden, 2003). Self-perceptions of personal initiative and supervisor rating of personal initiative and performance are reflective indicators of personal initiative. That is, variation in the preference to show high vs. low initiative in simulated work situations is reflected in the variation in self-concepts and supervisors' perceptions of personal initiative and performance (see Figure 3). Bollen (1984) refers to such models as multiple-indicator multiple-cause models (MIMIC). The fit of the model was $\chi^2 = 8.78$ ($df = 5$), $CFI = .96$ and $RMSEA = .078$. The factor loadings for the reflective indicators were .50 for self-rated personal initiative, .85 for supervisor rated personal initiative and .61 for overall performance³. For all further analysis we used the average across the twelve SJT-PI items as an unweighted manifest variable (see method section).

 Insert Figure 3 about here

Table 1 presents means, standard deviations and correlations of all variables. With respect to the validity of the SJT-PI the relationship between the SJT-PI and supervisors' personal initiative ratings was .48 ($p < .01$). The relationship between the SJT-PI and supervisors' overall performance ratings was .37 ($p < .01$) confirming hypotheses 1 and 2 and supporting the criterion validity of the SJT-PI. As expected in hypothesis 3a we found a moderate positive relationship ($r = .29, p < .01$) between the SJT-PI and self-ratings of personal initiative. Thus, both measures of personal initiative converged only partly.

 Insert Table 1 about here

We used path analysis to test hypotheses 3b – 4b simultaneously. The results of the path analysis are presented in Figure 4. All fit indexes indicate good fit for the full mediation model ($\chi^2 = 6,7$ with $df = 5$, $CFI = .99$ and $RMSEA = .053$). First, we note that the two operationalizations of personal initiative explain unique variance in both supervisor rated personal initiative and performance supporting the notion that both measures are complementary (Hypothesis 3b). They tap different information about respondents' level of personal initiative. The psychological orientations generalized self-efficacy and felt responsibility in turn each explain unique variance in respondents' situated preferences for personal initiative and self-ratings of personal initiative (Hypothesis 4a). Descriptively the relationships between self-ratings of personal initiative and psychological orientations are higher than the relationships between the

SJT-PI and psychological orientations. Self-efficacy and felt responsibility are also bivariately related to supervisors' personal initiative and performance ratings (see Table 1). These relationships are fully mediated by situated preferences for personal initiative and self-ratings of personal initiative (Hypothesis 4b)⁴.

Insert Figure 4 about here

Hypotheses 5a and 5b pertain to the discriminant validity (Campbell & Fiske, 1959) of the SJT-PI relative to the organizational citizenship behavior dimensions helping and conscientiousness. As can be seen in Table 1 there was a significant positive relationship of the SJT-PI with supervisor ratings of helping behavior ($r = .28, p < .02$) and no relationship with self-reported helping behavior ($r = .07, p = .43$). To test hypothesis 5a we compared the magnitude of the relationships with the formula proposed by Steiger (1980). In support of the hypothesis the relationship between the SJT-PI and supervisors' ratings of helping behavior was significantly smaller ($p = .03$) than the relationship between the SJT-PI and supervisors' personal initiative ratings. With respect to conscientiousness the SJT-PI was unrelated to self-ratings ($r = .07, p = .43$) and supervisor ratings ($r = .18, p = .12$). The (non-significant) relationship between the SJT-PI and supervisors' ratings of conscientiousness was significantly smaller ($p < .01$) than the relationship between the SJT-PI and supervisors' personal initiative ratings lending support to hypothesis 5b. Supervisor ratings of the three constructs helping, conscientiousness and personal initiative were significantly interrelated ($r = .50 - .53, p < .01$) indicating a halo effect or a higher order performance factor (Viswesvaran & Ones, 2000).

Discussion

This work was built on the observation of a misfit between the concept of personal initiative as situated action and the way it is measured in current organizational research. We proposed to improve measurement of personal initiative by not abstracting from the context in which initiative actions occur. Results of a validation study and qualitative data on item responses support our reasoning. A situational judgment test allows for a more situated way of measuring personal initiative. This implies that actions are in interaction with the situational context and that the specific situational context is important for what people do.

Results of the validation study supported the validity of the situational judgment test of personal initiative. Respondents' situated behavioral preferences were related highly to personal initiative as perceived by supervisors. Furthermore, the SJT-PI was related to supervisors' ratings of overall performance indicating that supervisors acknowledge the positive contribution of personal initiative to general performance. The resulting pattern of convergent and discriminant validities of the SJT-PI demonstrates that the measure assesses personal initiative but not other performance dimensions. The SJT-PI could be sufficiently differentiated from supervisor and employee ratings of helping and conscientiousness. Furthermore, the SJT-PI and the Likert scale of personal initiative were both related to self-efficacy and felt responsibility which have been discussed as antecedents of personal initiative.

As expected there was only a moderate positive relationship between the SJT-PI and the self-report Likert scale of personal initiative. Both measures tap different information about respondents' personal initiative. The Likert scale measures a broad self-concept. In contrast, the SJT-PI measures situated preferences for personal initiative in various simulated work situations. A qualitative analysis on how respondents answer both measures supports this interpretation. In

the case of the SJT-PI respondents immersed themselves into the scenarios and revealed their relative preference for different actions while taking into account the situational context. In the case of the Likert scale respondents either directly related to their self-concept or inferred their answer from past behavioral examples. Both operationalizations of personal initiative predicted independent variance in supervisor ratings of personal initiative and performance. This type of result is not unusual in the literature and has appeared in other areas of research, for example in the area of achievement motives (Brunstein, 2004; Bing, 2007). Both Likert-type as well as TAT responses both show to have validity but are little related with each other (Spangler, 1992).

We conceptualized specific instances of personal initiative and their simulation in the SJT-PI as formative indicators of the personal initiative construct. Personal initiative is a performance concept that describes an observable aspect of work performance rather than a psychological attribute of individuals (Frese & Fay, 2001). The overall level of employees' personal initiative as it is reflected in self and supervisor ratings is a function of the relevant behaviors employees perform across different situations at work. Specific instances of personal initiative can have the same or different antecedents and do not necessarily have to be related. Measuring personal initiative across very different situations in a low fidelity simulation ensures that the construct is not represented too narrowly and that different scenarios add incremental validity to the prediction of an external criterion (MacKenzie et al., 2005).

A challenge for research and application of SJTs is the estimation of reliability and the related issue of adequate number of scenarios of a test. The requirements of unidimensionality and essential tau equivalence for the estimation of internal consistency are not assumed in a formative approach to measurement. Internal consistency does thus not provide a guideline for selecting scenarios and determining the number of scenarios required to estimate respondents'

overall level of personal initiative. Test-retest reliability has been proposed as an alternative estimate of reliability (Motowidlo et al., 1990). Concerning the test-retest reliability of the SJT-PI we estimated test-retest reliability with a subset of participants and items and found adequate stability of responses over time. Past research has also found sufficient stability of SJTs over time to use them in selection (e.g. Ployhart & Ehrhart, 2003).

A critical issue for the SJT-PI is to include a representative sample of initiative actions as scenarios. In contrast to latent variables with reflective indicators, constructs with formative indicators are defined by their indicators. That is, the meaning of the SJT-PI is dependent on the scenarios included in the test. Changing or replacing the scenarios can change the meaning of the construct. Therefore, it was essential to include a representative sample of instances of personal initiative in the SJT-PI. Since only a sample of situated preferences is used to assess respondents overall preference for personal initiative, there remains measurement error associated with the construct (see Figure 3). In the absence of a theoretical taxonomy of situated initiative actions, we used an empirical strategy to obtain a heterogeneous sample of situated initiative actions. Employees in a variety of jobs reported instances of personal initiative that had occurred. A limitation of this approach is that the sample of 12 items may not comprehensively reflect the domain of personal initiative. A different sample of scenarios may provide a test that is only moderately related to the SJT-PI but still predicts behavioral criteria of personal initiative. However, this does not question the usefulness of the current SJT-PI but indicates potential for improvement. Higher validities may be achieved if scenarios are added to the test that address different kinds of situated initiative actions. We think further development of the SJT-PI should not only be based on empirical evidence but also on a theoretical taxonomy of situations. Our exploratory analysis on situational factors underlying item responses points to a similar direction

as recent research by Griffin et al. (2007). They found initiative actions of individuals in different roles to be distinct phenomena. Individual task proactivity, team member proactivity and organization member proactivity were found to be three different factors underlying both self- and supervisor ratings.

Concerning future application of the SJT-PI we are confident that the test can be applied in office environments. Although the SJT-PI was validated with a sample of bank employees, the test did not contain any items that were specific for working in a bank. The test was based on critical incidents collected in a wide variety of jobs and was constructed before the validation sample was determined. Future research needs to examine if the test can also be applied in work settings other than office environments. While the SJT-PI can be used for research on personal initiative, we do not recommend application of the test for personnel selection pending further research. In particular the test's incremental validity over established selection methods and its susceptibility to faking need to be examined.

Recommendations for the Development of Situational Judgment Tests

For the development of SJTs measuring constructs we recommend different methodological approaches depending on the objective. If an in-depth understanding of the situational facets of a behavioral construct is the goal, reliable measurement of each facet is crucial. This applies to research on differential relationships of situational facets to antecedents and outcomes and to research addressing questions of a person by situation interaction using SJTs. To achieve reliable measurement, situational facets should be each measured with multiple homogeneous scenarios as reflective indicators. In such a model, the overall level of a construct is determined by the situational facets as latent formative indicators. Each situational facet is measured by multiple SJT items as reflective indicators (MacKenzie et al., 2005). By this means

a limitation of formative measurement models (Howell et al., 2007) can be addressed: The meaning of a latent variable with formative indicators is dependent on the criteria included in a model. If situational facets are measured with reflective indicators their meaning is not dependent on the outcome criteria. Only the latent variable composite construct defined by the situational facets has to be interpreted relative to the criteria. The weight with which each situational facet contributes in defining the composite construct changes dependent on the criteria and therefore the meaning of the composite construct also changes. For instance, if we had used career success as an alternative criterion of personal initiative, scenarios that are directly relevant to career success should have contributed more strongly to the latent variable they define. The meaning of personal initiative in this model would change accordingly. Concerning reliability, future attempts to develop SJTs with situational facets can use stratified alpha as a reliability estimate. However, stratified alpha requires multiple essentially tau equivalent items as reflective indicators of each stratum, i.e. situational factor or facet of a test (Rae 2007).

The complexity of SJT-items and the sensitivity of responses to the contextual information provided in the items may pose constraints in developing multiple SJT-items that load highly on one situational facet without creating highly redundant items and a test that is too time consuming for applied purposes. Researchers and practitioners who are interested in a composite construct and not in its facets may therefore use a different approach: Rather than developing homogeneous scenarios, items should be a heterogeneous and representative sample of the construct domain. By this means, situational judgments tests can capture the breadth of a construct without consisting of too many scenarios to be efficiently applied. As internal consistency is not the primary criterion for selecting items, other information needs to be used

for selecting items. We used the level of agreement by subject matter experts about the extent to which response options represent different levels of the construct. If a formative model is used, choosing the right criterion measures is essential. The meaning of the latent variable with formative indicators is dependent on the criteria. To determine the latent construct we recommend using criteria that are direct manifestations of the construct such as self ratings, other ratings and behavioral outcomes. If situational facets are examined, narrow criteria for each situational facet can provide additional information on a test's validity. As an alternative to modeling a formative model, the unweighted average can be used as a predefined measure of a construct. The unweighted average is not dependent on criterion measures although this might result in information loss (Howell et al. 2007).

Conclusion

This study has demonstrated that an SJT can be developed to measure a specific behavioral construct - personal initiative. The good criterion validity of this measure that is difficult to find for self-report measures supports the notion that this situated approach to measurement is promising. We think that the SJT approach is particularly useful for behavioral constructs that already have a clear and specific theoretical definition. In contrast to many Likert-type scales, SJTs do not base measurement on decontextualized and generalized statements but on specific behavioral examples. We argue that this prevents ambiguity with respect to the meaning respondents attach to items. Research is needed that applies different SJTs measuring independent constructs in the same sample. This would clarify to what extent there is a general SJT method factor that confounds the constructs that are measured.

References

- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50, 248-287.
- Bing, M. N., LeBreton, J. M., Davison, H., Migetz, D. Z., & James, L. R. (2007). Integrating implicit and explicit social cognitions for enhanced personality assessment: A general framework for choosing measurement and statistical methods. *Organizational Research Methods*, 10(2), 346-389.
- Bolino, M. C., & Turnley, W. H. (2005). The Personal Costs of Citizenship Behavior: The Relationship Between Individual Initiative and Role Overload, Job Stress, and Work-Family Conflict. *Journal of Applied Psychology*, 90(4), 740-748.
- Bollen, K. (1984). Multiple indicator: Internal consistency or no necessary relationship? *Quality and Quantity*, 18, 377-385.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305-314.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review Vol 110(2) Apr 2003*, 203-219.
- Brunstein, J. C., & Maier, G. W. (2005). Implicit and Self-Attributed Motives to Achieve: Two Separate but Interacting Needs. *Journal of Personality and Social Psychology*, 89(2), 205-222.
- Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: A user's guide. *Organizational Research Methods*, 5(2), 159-172.

- Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On Average Deviation Indices for Estimating Interrater Agreement. *Organizational Research Methods*, 2(1), 49-68.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Costall, A. & Leudar, I. (1996). *Situating Action I: Truth in the Situation*. *Ecological Psychology*, 8(2) 101-110.
- Crant, J. M. (2000). Proactive behavior in organizations. *Journal of Management*, 26, 435-462.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155-174.
- Ericsson, K. A., and Simon, H. A., 1993. *Protocol analysis: Verbal reports as data*. The MIT Press, Cambridge, MA.
- Farh, J.-I., Podsakoff, P. M., & Organ, D. W. (1990). Accounting for organizational citizenship behavior: Leader fairness and task scope versus satisfaction. *Journal of Management*, 16(4), 705-721.
- Fay, D., & Frese, M. (2001). The concept of personal initiative: An overview of validity studies. *Human Performance*, 14(1), 97-124.
- Frese, M., & Fay, D. (2001). Personal initiative: an active performance concept for work in the 21st Century. *Research in Organizational Behavior*, 23, 133-187.
- Frese, M., Fay, D., Hilburger, T., & Leng, K. (1997). The concept of personal initiative: Operationalization, reliability and validity of two German samples. *Journal of Occupational & Organizational Psychology*, 70, 139-161.

- Frese, M., Teng, E., & Wijnen, C. J. D. (1999). Helping to improve suggestion systems: Predictors of making suggestions in companies. *Journal of Organizational Behavior*, 20, 1139-1155.
- Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H. C. Triandis & M. D. Dunnette (Eds.), *Handbook of industrial and organizational psychology, Vol. 4 (2nd ed.)*. (pp. 271-340): Consulting Psychologists Press, Inc.
- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54(7), 493-503.
- Griffin, M. A., Neal, A. S., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal*, 50(2), 327-347.
- Hacker, W. (2003). Action Regulation Theory: A practical tool for the design of modern work processes? *European Journal of Work and Organizational Psychology*, 12 (2), 105–130
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205-218.
- König, C., Steinmetz H., Frese, M., Rauch, A., & Wang Z-M., (2007). Scenario-based scales measuring cultural orientations of business owners. *Journal of Evolutionary Economics*, 17, 211–239.
- Koop, S., De Reu, T., & Frese, M. (2000). Sociodemographic factors, entrepreneurial orientation, personal initiative, and environmental problems in Uganda. In M. Frese (Eds.), *Success and failure of microbusiness owners in Africa: A psychological approach* (pp. 55-76), Westport, Ct.: Quorum.

- Krauss, S. I., Frese, M., Friedrich, C. (2007). *Psychological action strategy characteristics, entrepreneurial orientation, and business performance: A longitudinal analysis among Zimbabwean small business owners*. University of Giessen, Department of Psychology: in preparation.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R., & Smith, D. E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance*, 13(4), 323-353.
- Latham, G. P., & Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology*, 69, 569-573.
- Latham, G. P., & Skarlicki, D. P. (1995). Criterion-related validity of the situational and patterned behavior description interviews with organizational citizenship behavior. *Human Performance*, 8, 67-80.
- Law, K. S., & Wong, C. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction construct. *Journal of Management*, 25(2), 143-160.
- MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, 90(4), 710-730.
- McClelland, D. C. (1987). *Human motivation*. Cambridge, England: Cambridge Univ. Press.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730-740.

- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection & Assessment*, 9, 103-113.
- Mischel, W. (2004). Toward an integrative science of the person., *Annual Review of Psychology*, 55, 1-22.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246-268.
- Morrison, E. W., & Phelps, C. C. (1999). Taking charge at work: Extrarole efforts to initiate workplace change. *Academy of Management Journal*, 42, 403-419.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75(6), 640-647.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91(4), 749-761.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475-480.
- Nardi, B. A. (1996). *Studying context: A comparison of activity theory, situated action models, and distributed cognition*, chapter 4, pages 35-52. The MIT Press, Cambridge, MA.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, 10(2), 85-97.
- Organ, D. W., & Konovsky, M. (1989). Cognitive versus affective determinants of organizational citizenship behavior. *Journal of Applied Psychology*, 74(1), 157-164.

- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a Biodata Measure and Situational Judgment Inventory as Predictors of College Student Performance. *Journal of Applied Psychology, 89*(2), 187-207.
- Parker, S. K., Williams, H. M., & Turner, N. (2006). Modeling the antecedents of proactive behavior at work. *Journal of Applied Psychology, 91*(3), 636-652.
- Peeters, H. and Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement, 65*(1), 70-89.
- Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods, 2*(4), 329–344.
- Ployhart, R. E. & Ehrhart, M. G., (2003). Be careful what you ask for: effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11*(1), 1-16.
- Ployhart, R. E., & Ryan, A. M. (2000). *A construct-oriented approach for developing situational judgement tests in a service context*. Unpublished manuscript.
- Rae, G. (2007). A note on using stratified alpha to estimate the composite reliability of a test composed of interrelated nonhomogeneous items. *Psychological Methods, 12*(2), 177-184.
- Schmidt, C. (2004). *Der Einfluss soziokultureller Kontexte auf das Arbeitsverhalten von Mitarbeitern*. University of Giessen: Unpublished manuscript.
- Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. A. Weekley and R. E. Ployhart (Eds.), *Situational judgment tests: Theory, and application* (pp. 135-156). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

- Sherer, M. (1982). The Self-efficacy Scale: Construction and validation. *Psychological Reports*, 51(2), 663-671.
- Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, 68, 653-663.
- Speier, C., & Frese, M. (1997). Generalized self-efficacy as a mediator and moderator between control and complexity at work and personal initiative: A longitudinal field study in East Germany. *Human Performance*, 10, 171-192.
- Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, 112, 140-154.
- Spencer, L. M. J., & Spencer, S. M. (1993). *Competence and work: Models for superior performance*. New York: Wiley.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251.
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communications*. Cambridge, UK: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5, 207-232.
- Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management Journal*, 41, 108-119.
- Viswesvaran, C., & Ones, D. S. 2000. Perspectives on models of job performance. *International Journal of Selection and Assessment*, 6, 216-226.
- Weekley, J. A., & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, 18(1), 81-104.

Zempel, J. (1999). Selbstaendigkeit in den neuen Bundeslaendern: Praediktoren, Erfolgsfaktoren und Folgen - Ergebnisse einer Laengsschnittuntersuchung. In K. Moser & B. Batinic & J. Zempel (Eds.), *Unternehmerisch erfolgreiches Handeln* (pp. 69-91). Goettingen, Germany: Verlag fuer Angewandte Psychologie.

Footnotes

¹ In the case of a construct with reflective indicators only the common variance shared by all indicators is considered true variance of the construct (MacKenzie et al., 2005). Variances specific to one situation, covariances shared by some situations and random variance are treated as error variance. In contrast, under the composite view only random variances are considered error variance. Variances specific to one situation or covariances shared by some situations are part of the true variance of the composite construct (Law & Wong, 1999).

² Alpha coefficients of SJTs are usually low, unless very large numbers of items are used (Schmitt & Chan, 2006). In their meta-analysis McDaniel et al. (2001) reported an artifact distribution of SJT alpha values ranging from .43 to .94 with a mean value of .60. In contrast to SJTs applied in personnel selection we used a much smaller number of items to have an efficient means of measurement for personal initiative at work. Cronbach's alpha of the 12 item SJT-PI was .61.

³ In spite of the problems present if items are not parceled (non-normal distribution, multicollinearity and number of indicators), we tested a model without parceling items to examine if results would change. After removing one item that has a direct relationship with supervisors' personal initiative ratings, the model showed good fit to the data ($\chi^2 = 16,3$ with $df = 21$; $CFI = 1.00$; $RMSEA = .000$). The three reflective indicators had loadings of .61 for self-rated personal initiative, .71 for personal initiative as perceived by supervisors and .47 for overall performance ratings.

⁴ To test whether the SJT-PI explained incremental variance in supervisor ratings, we conducted two multiple hierarchical regression analyses with the criteria overall performance

and supervisor ratings of personal initiative. The order predictors were entered was based on their theoretical proximity to the dependent variables. After entering the control variables age, gender and job position in step 1, we added self-efficacy and felt responsibility in step 2. In step 3 the self-ratings of personal initiative were added and in step 4 the SJT-PI was added to the model. After controlling for all other variables, the SJT-PI explained additional 9% of variance ($p < .01$) in supervisor ratings of personal initiative and 4% ($p = .04$) of variance in supervisors' performance ratings.

Table 1

Means, Standard Deviations and Correlations

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10
<i>Self-ratings^a</i>												
1. SJT-PI	.43	.51	(-)									
2. Personal Initiative	3.79	.49	.29**	(.80)								
3. Conscientiousness	3.57	.83	.07	.16 [†]	(-)							
4. Helping behavior	4.15	.54	.07	.25**	.22*	(.80)						
5. Self-Efficacy	4.51	.40	.30**	.45**	.07	.11	(.79)					
6. Felt Responsibility	3.68	.63	.28**	.47**	.33**	.28**	.25**	(.77)				
<i>Supervisor Ratings^b</i>												
7. Performance	4.82	1.26	.37**	.27*	.12	.02	.28*	.36**	(.96)			
8. Personal Initiative	3.46	.75	.48**	.40**	.11	.00	.27*	.36**	.72**	(.93)		
9. Conscientiousness	3.82	.79	.18	.32**	.03	-.01	.12	.14	.49**	.53**	(.81)	
10. Helping behavior	3.88	.67	.28*	.26*	.06	.05	.09	.27*	.63**	.50**	.51**	(.89)

Note. Reliability estimates are presented in the diagonal (test-retest reliability for the SJT-PI, Cronbach's alpha for all other scales)

^a *N* = 126. ^b *N* = 77.

[†] *p* < 0.10. **p* < 0.05. ***p* < 0.01.

Figure Captions

Figure 1. Example Item of the SJT-PI.

Figure 2. Theoretical Model with Self-Reported Personal Initiative and Situated Preferences for Personal Initiative as Mediators.

Figure 3. MIMIC-Model with SJT-PI Item Parcels as Formative Indicators.

Figure 4. Test of a Path Model with Self-Reported Personal Initiative and Situated Preferences for Personal Initiative as Mediators.

Figure 1

12. You are under enormous pressure to accomplish your tasks on time. Yesterday, new trainees started in your department. They are unfamiliar with the workflow in your department. You have to interrupt your work to answer trainees' questions and to correct their mistakes. You are expected to do both, to finish your work on time and to take care of the trainees. What would you do?

least likely		most likely
<input type="checkbox"/>	I tell the trainees that I am available after work to answer their questions.	<input type="checkbox"/>
<input type="checkbox"/>	I openly say that I cannot take care of the trainees and work for better initial training of the trainees.	<input type="checkbox"/>
<input type="checkbox"/>	I send the trainees to my colleagues when they have questions.	<input type="checkbox"/>
<input type="checkbox"/>	I try to get by without becoming stressed and worn out.	<input type="checkbox"/>

Figure 2

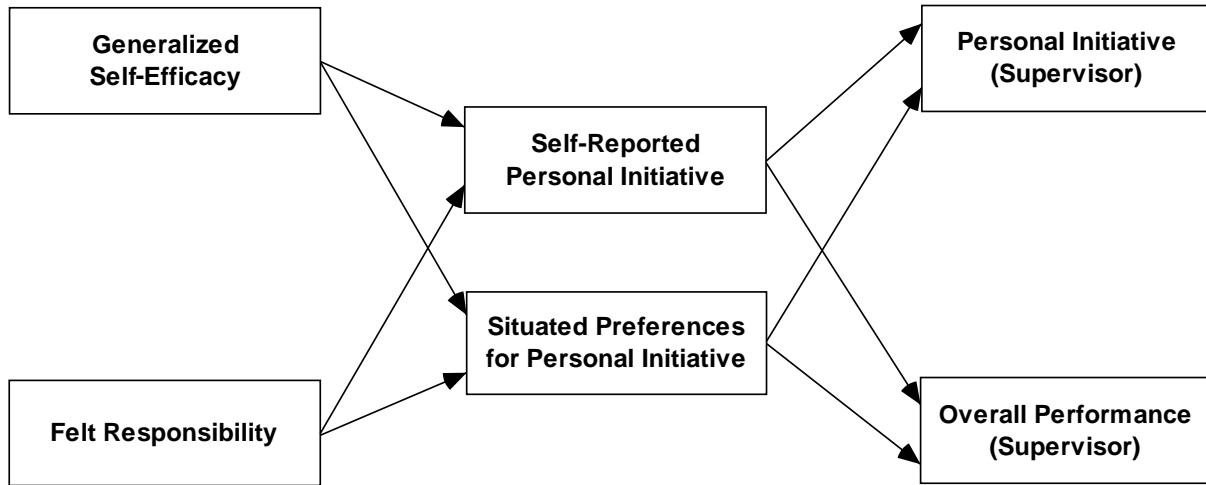
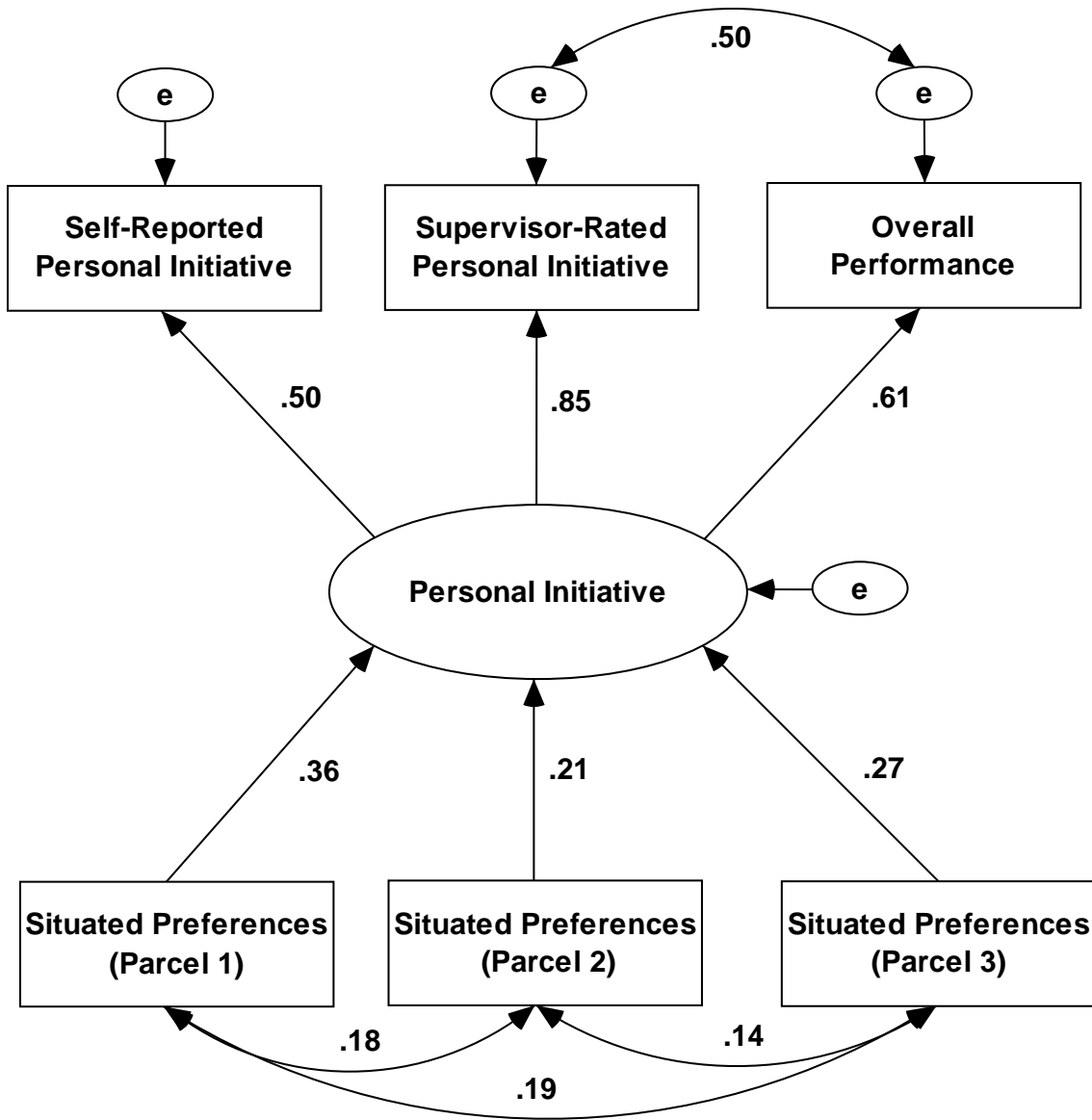


Figure 3



Note that the composite construct has an associated error term. Representing the error at the construct level takes into account that aspects of the construct domain are not represented in the indicators. The error term reflects that the construct has a surplus meaning and is more than an empirical combination of measures (MacKenzie et al., 2005).

Figure 4

