

The validation of a video-based situational judgment test for the selection of call center employees



Maartje Harlaar

Student number: 1930265

Supervisor: Dr. J. K. Oostrom

2nd Supervisor: Dr. R. E. de Vries

Master Thesis Psychology: Work and Organizational Psychology

August, 2013

Vrije Universiteit Amsterdam

Abstract

In this study a construct-driven video-based situational judgment test (SJT) was developed for the selection of call center employees. The SJT was intended to measure six communication styles which were derived from an existing communication model. Study 1 investigated the construct validity of the SJT with data of 147 non call center employees who completed the SJT and the CSI. Contrary to our expectation, results showed no significant correlations between the scores on the domain-level scales of the SJT and the scores on the corresponding domain-level scales of the CSI. Therefore, there was no evidence found for the construct validity of the SJT. Study 2 involved a sample of 146 call center employees working in an inbound call center. The results showed no significant correlations between the overall SJT score and job performance of the call center employees. In addition, the SJT did not incrementally predict job performance in call centers beyond personality and cognitive ability. The results in Study 2 were probably influenced by the way the performance of the employees was rated. Furthermore, call center employees scored significantly higher on the SJT than non call center employees. This indicates that the SJT is able to distinguish between experienced employees and non experienced employees. Overall, the SJT seems not appropriate to measure communication. Given the limitations of the job performance measures, further research is needed to examine the predictive validity of the SJT.

1. Introduction

Companies in diverse business sectors commit a great deal of time and resources on customer satisfaction. Delivering superior service and ensuring higher customer satisfaction have become strategic necessities for companies to survive in competitive business environments (Jaiswal, 2008). Customer call centers have emerged as an important tool for providing higher customer satisfaction (Aksin & Harker, 1999). Call centers allow a company to build, maintain, and manage customer relationships by solving problems and resolving complaints quickly, providing information, answering questions, and being available usually 24 hours a day, seven days a week, 365 day of the year (Prahabkar, Sheehan, & Coppett, 1997). Because it is usually the first contact point of an organization with its customers, the impressions on the total service quality are often made from the call center interactions (Ma, Kim, & Rothrock, 2011).

In recent years, call centers have grown rapidly in volume and popularity over the world (Ramseook-Munhurrun, Naidoo, & Lukea-Bhiwajee, 2009). This rapid rise of call centers has been accompanied by a significant challenge in attracting and retaining employees. The estimated average turnover per year is between 35 and 50 percent (Sawyerr, Srinivas, & Wang, 2009). To succeed as an organization, employers need to recruit applicants who perform well on the job (McCulloch & Turban, 2007). Attracting talented employees increases the reputation and image of the call center organization (Smith, 2001). So far, however, researchers have paid little attention to call center recruitment and selection. This is surprising given that call centers are part of a wider trend towards increasing service work in the economy.

Generally, the first step to attract qualified employees is through hiring practices. Additional time and resources should be spent at the beginning of the employment process (Bloomquist & Kleiner, 2000). Given the above information it would be useful for organizations to have a selection tool that is able to predict job performance. Because most applicants will be unfamiliar with actual call center activities, the selection tool should also be intended to inform the applicant of what the job would look like. A realistic job preview (RJP), which presents both the positive and negative elements of the job and the work environment, will provide a strong incentive for people who are going to fit,

and forewarn those who are not going to fit anyway (Smith, 2001). Philips (1998) found that RJPs were related to higher job performance.

Several studies have demonstrated that situational judgment tests (SJTs) provide a realistic job preview (Weekley & Ployhart, 2006) and are able to predict job performance (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Chan & Schmitt, 2002). SJTs confront applicants with written or video-based work-related situations and ask them to indicate how they would react by choosing an alternative from a list of responses (Christian, Edwards, & Bradley, 2010; McDaniel, Hartman, Whetzel, & Grubb, 2007).

The purpose of the present study is to develop and validate a video-based SJT for the selection of call center employees. Before the development of the SJT will be described, there will be an overview presented of the literature on call centers and SJTs. After that, several hypotheses about the construct validity, predictive validity, and incremental validity of the SJT are proposed.

1.1. Call centers

In a call center calls are placed or received for the purpose of sales, marketing, customer service, telemarketing, technical support, and other business activities (Bodin & Dawson, 1999). The way call centers get in contact with customers may differ. Whereas inbound call centers have a more passive role (i.e., being called up exclusively by customers with questions or complaints concerning a product), outbound call centers actively engage in calling people (e.g., telemarketing call centers). However, there are also call centers with both inbound and outbound activities (Zapf, Isic, Bechtoldt, & Blau, 2003).

Although customers' perceptions of the company are determined by the quality of the interaction with its frontline employees in call centers (Mattila & Mount, 2003; Peccei & Rosenthal, 1997), several studies highlight the fact that minimal attention has been paid to developing criteria for the selection of call center employees (Bain & Taylor, 2000; Hutchinson, Purcell, & Kinnie, 2000). It is important to develop criteria because selection of the right call center employees is critical to call center success (Nicholls, Viviers, & Visser, 2009). To be able to develop criteria that can help identify successful call center employees, we have to first understand which individual differences are related

to job performance in call centers (Sawyer et al., 2009).

In general, cognitive ability and personality are the two most commonly studied individual difference predictors of job performance (Chan & Schmitt, 2002). Sawyer et al. (2009) examined the relationship between personality traits and service performance in the context of a call center environment. Their results showed a negative relationship between openness to experience and service performance. There were no significant relationships between service performance and the other personality traits. Although several studies found strong relationships between the Big Five personality traits and performance (Mount & Barrick, 1998; Barrick & Mount, 1991), the results of Sawyer et al. indicate that these may not readily apply to the call center environment.

Cognitive ability has been shown to have a corrected mean correlation of .45 for predicting job performance across a wide variety of occupations (Hunter & Hunter, 1984). However, research has paid little attention to the relation of cognitive ability and job performance in a call center context. Despite its high predictive validity, measures of cognitive ability have a few important disadvantages as selection instrument. First, cognitive ability does not predict job performance on low complexity tasks as well as job performance on more complex tasks. Second, cognitive ability tests have been shown to produce adverse impact (Hunter & Hunter, 1984). Adverse impact occurs in an employment decision when minority group members are selected at a substantially lower rate than members of the majority group (Morris & Lobsenz, 2000).

A potentially valuable predictor of job performance of call center employees, which has received little attention in the call center literature so far, is communication style. Of several factors that determine customer satisfaction, communication style is one of considerable importance because of its central role in connecting employees and customers (Mohr & Navin, 1990) and in establishing customer trust and satisfaction (Ring & Van de Ven, 1994). Although the literature suggests that the communication style of an employee is likely to affect the quality of the service encounter by influencing the customer's impression of the organization and the service firm (Webster & Sundaram, 2009), there is a lack of research on the relationship between communication styles of call center employees and their job performance.

Recently, a multiphase lexical study was conducted to uncover the key dimensions of communication styles (De Vries, Bakker-Pieper, Altin Siberg, Van Gameraen, & Vlug, 2009). Based on this research, De Vries, Bakker-Pieper, Konings, and Schouten (2011) introduced the Communication Style Inventory (CSI) which measures six behavioral communication style dimensions: Expressiveness (X), Preciseness (P), Verbal Aggressiveness (VA), Questioningness (Q), Emotionality (E) and Impression Manipulativeness (IM). Personality and communication styles are interlinked (McCroskey & Beatty, 2000) because someone's communication style may be viewed as an expression of one's personality. De Vries et al. (2011) found strong evidence for the relationships between the CSI domain-level scales and two well-known personality measures, namely the HEXACO-PI-R (Asthon & Lee, 2008; De Vries, Asthon, & Lee, 2009; Lee & Ashton, 2004) and the NEO-PI-R (Costa & McCrae, 1992; Hoekstra, Ormel, & De Fruyt, 1996). Bakker-Pieper and De Vries (2013) examined the relationship between the CSI and leader outcomes. Their results showed that expressive and precise communication styles of a leader were related to leader outcomes. Expressive and precise communication styles also showed incremental validity over the personality traits of extraversion and conscientiousness in the prediction of leader outcomes. There are no other studies that have examined the relation between the CSI and job performance. In the present study the CSI will be used as a basis to develop the SJT for call center employees. Using an SJT instead of a self-report questionnaire like the CSI for the selection of call center employees has several benefits. SJTs are more face valid and job related, provide samples of job behavior, and provide information about the job to the applicants (Bauer & Truxillo, 2006). Several other benefits of the SJT will be discussed in the following section.

1.2. Situational Judgment Tests

SJT's are a popular and useful method for identifying qualified applicants (Weekly, Ployhart, & Harold, 2004). They present applicants with work-related situations and ask them how they would or should respond to each situation. The situations and reactions are usually presented in written, verbal, video-based, or computer-based formats (e.g., Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt-Harvey, 2001; Motowidlo, Dunnette, & Carter, 1990). The situations are created from either

critical incidents or other job content material (Ployhart & Ehrhart, 2003). SJTs may be classified as job simulations. Simulations are based on the assumption that one can predict how well an individual may perform on a job, based on how the individual performs on a simulation of the job (McDaniel & Nguyen, 2001).

The increased popularity of SJTs is undoubtedly due to research showing these tests to have a number of positive features. SJTs have moderate criterion-related validity, have high face validity, and show smaller racial and gender subgroup differences than cognitive ability tests (Chan & Schmitt, 2002; McDaniel & Nguyen, 2001; Weekley et al., 2004). SJTs also have incremental criterion-related validity over other common selection methods such of cognitive ability, job knowledge, job experience, and conscientiousness (Clevenger et al., 2001). New technology has made the development of SJTs based on video material possible. The video based SJT appears to have several advantages compared to the paper-and-pencil SJT, such as a higher criterion-related validity (Lievens & Sackett, 2006), less adverse impact, and higher realism leading to more reliable respondent reactions (Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000).

Despite the widespread use of SJTs, the literature reveals that test developers and researchers often give little attention to the constructs measured by SJTs and tend to report results based on overall (or composite) SJT scores. But to understand how and why SJTs work in selection context, it is critical to identify the constructs assessed with SJTs (Christian et al., 2010). The search for underlying constructs should enable one to better justify the use of selection procedures and their outcomes to the candidates and the organization. A so-called construct-driven approach will significantly advance our understanding of the criterion (the performance domain) that these instruments aim to predict (Lievens, Van Dam, & Anderson, 2002). This study is innovative in that a construct-driven approach will be used to develop a video-based SJT for communication styles. In two studies, its construct validity, predictive validity, and incremental validity will be examined.

1.3. Hypothesis of Study 1

Traditionally, the primary goal in personnel selection is to develop selection instruments that are able to predict candidates future work performance (Lievens et al., 2002). However, the goal

should not just be to show that a measure predicts job performance but also why that measure or construct predicts job performance (Arthur & Villado, 2008; Messick, 1995). Whereas meta-analytical research has shown SJTs to have good predictive validity, it is still unclear which constructs are exactly measured. To date, the general assumption has been that SJTs measure general cognitive ability, because substantial correlations ($r = .46$) with cognitive ability tests have been reported (McDaniel & Nguyen, 2001). Meta-analysis of the construct validity data showed that there are average correlations between SJTs and agreeableness, conscientiousness, emotional stability, extraversion and openness (McDaniel & Nguyen, 2001). However, the correlations between SJT scores and personality test scores often vary quite widely depending on the specific content of the SJT. As noted before, the SJT research could benefit from developing a more construct-driven approach. Therefore, in the present study a construct-driven approach to develop a video-based SJT will be used.

The video-based SJT is intended to measure the six communication styles of the CSI. The aim of Study 1 was to investigate the construct validity of the SJT. Because the SJT is constructed based on the communication styles of the CSI, we expect positive correlations between the SJT domain-level scales and the corresponding CSI domain-levels scales, such that SJT Expressiveness is most strongly related to CSI Expressiveness, SJT Preciseness is most strongly related to SJT Preciseness, and so on. Therefore, the hypothesis is:

Hypothesis 1: Scores on the domain-level scales of the SJT have a positive relationship with the domain-level scales of the CSI.

2. Method Study 1

2.1. Participants and Procedure

Participants were recruited from an Internet Panel. Inclusion criteria were (a) at least 18 years of age, (b) younger than 55 years of age, (c) graduated at level 3 or 4 of Intermediate Vocational education. These criteria were used because this sample had to be representative for employees who are working in a call center. Demographic characteristics of call center employees were used to develop these criteria. The sample consisted of 147 non call center employees. Age was divided in the following categories: 18-25, 26-35, 36-45 and 46-55. The majority fell into the categories 36 – 45

(36.7%) and 46- 55 (36.7%). Ninety-nine participants were female (67.3%) and 48 participants were male (32.7%). Their level of education was Intermediate Vocational. In the Netherlands there are four levels of Intermediate Vocational education, 44 participants had level 3 (46%) and 101 participants had level 4 (68%). Beside age, sex and educational level, we asked no other background information. The participants completed the CSI and after that they completed the SJT. Participants received money as a reward for their participation. This reward was a standard amount that members of the panel received for completing a questionnaire. The SJT items were presented in a game-like portal that also contained other types of tests. These included a multitasking test, a language test, and a typing test. This study focuses on the SJT items.

2.2. Measures

2.2.1. SJT

In this study a construct-based SJT was developed to measure communication styles. Development procedures of Ployhart, Porr, and Ryan (2004) were followed to develop the SJT and to assess its construct validity. In their study, a SJT was designed to assess the three personality constructs most relevant for customer service jobs. Results supported the construct validity of the SJT, which meant that the SJT items reflected personality traits.

The first step was to define the performance domain by reviewing the relevant call center literature. From the literature it was suggested that communication could be an important predictor of performance in a call center (Callaghan & Thompson, 2002; Webster & Sundaram, 2009). A six-dimensional behavioral model of communication styles, the so-called CSI (De Vries et al., 2011) was found to be a valid measure of communication styles. Therefore, it was decided to use the CSI as a basis for the development of the SJT items. The next step was to identify situations that would result in maximal variability in the demonstration of these communication styles such that individual differences in call centers could be manifested. Four call center experts provided information about critical incidents in the call center. Three of the experts were trainers who give training to call center employees during their first weeks at the call center. The other expert was a manager of the HR-department. They were asked to think of situations that were (1) frequent and common, (2) difficult or

challenging to handle, and (3) important. Situations that are psychologically demanding for a particular trait should be most likely to lead to individual differences in manifestations of that trait (Shoda, Mischel, & Wright, 1993). The experts provided situations that frequently occur during a working day of call center employees. An example situation is: 'It is typical in call centers to handle calls under time pressure in combination with unfriendly or angry customers' or 'Some customers talk too much during a call, then a call center employee have to interrupt and lead the conversation in a good way'. This input was used to create situations for the items. In the next step specific communication styles were linked to each situation. For example, a situation in which a customer communicated very unfriendly, was assigned to the 'verbal aggressiveness' communication style. The fourth step involved constructing the behavioral response options. We adopted a different method than typically is used to develop SJTs. Specifically; the response options were created to reflect a range of the specific communication style within a given situation. This should help to make the items more likely to assess the intended communication constructs. Following the example above, the response options for the 'verbal aggressiveness' situation reflected a range of this communication style:

- A. I will write it down and keep it in mind next time this situation occurs.
- B. I will consider you remarks.
- C. Maybe you should take a look at yourself.
- D. If you think I'm wrong, you can hang up the phone.

Four response options were constructed for each situation. This procedure resulted in 24 SJT items (a situation including its four response options is labeled 'item'). Four items were constructed for each of the six communication styles. After this, the content of the 24 items was examined. Three subject matter experts (SMEs) unaffiliated with this study but familiar with the CSI rated, independently of one another, both the situations and response options. Two of the three SMEs (3 females) had a master degree in work and organizational psychology and one SME had a master degree in cognitive research. Their age ranged from 29 to 48 ($M = 39.00$, $SD = 9.54$). Each SME read the items and made two judgments. First, each item was judged for its relevance to call center contexts. These ratings were based on a five-point scale, with 1 = *completely irrelevant* and 5 = *extremely*

relevant. An overall mean score for the responses of the three SMEs was calculated. The relevance of the items was rated relatively high ($M = 4.03$, $SD = 0.79$). The one-way random effects intra-class correlation (ICC) for absolute agreement was .56. This indicates that they did not completely agree on the relevance of the items. Second, the SMEs were asked to indicate to which of the six communication styles each item belonged by writing down a communication style. The one-way random effects intra-class correlation (ICC) for absolute agreements was .78, which indicates that there was substantively agreement about the classification of the communication styles for each item. The first SME classified twenty one items correctly, the second SME classified twenty items correctly, and the third SME classified ten items correctly. Together, there were 8 items which they all correctly classified. Based on the feedback of the relevance and the classification from these three SMEs, 12 items were rewritten. Rewriting included changing one or more response options of an item.

Video clips were made for these 24 items. These videos were recorded with non-professional actors. Subsequently, the SJT items were tested in a pilot study (non call center employees; $N = 65$, 23.1% male, 76.9% female, call center employees: $N = 68$, 38.2% male, 61.8% female). Alpha coefficients of the SJT domain-level scales were: $\alpha = .08$ for Expressivity, $\alpha = -.12$ for Emotionality, $\alpha = .29$ for Precision, $\alpha = .29$ for Verbal Aggressiveness, $\alpha = -.04$ for Impression Manipulation, and $\alpha = .33$ for Discussion Willingness. Reviews of the research and theory related to SJTs identify the reliability of SJTs as problematic. This is due to the fact that most SJTs are multidimensional measures, and are often construct heterogeneous at the item level (Lievens, Peeters, & Schollaert, 2008; Whetzel & McDaniel, 2009; Weekly & Ployhart, 2006). Therefore, alpha coefficients are usually relatively low. The shape of the response distribution for each item was inspected by histograms. These histograms indicated that a few of the items had one very attractive option and therefore limited variance. In these cases one or more of the response options were rewritten to make them more or less attractive alternatives and to increase variance across options. In total there were 16 items rewritten after this pilot study.

Subsequently, a new video-based version of the test was developed. The scenarios were videotaped in a professional manner with professional actors. The actor was playing a customer and

talked directly into the camera as if he or she talked to the participant. After the customer asked a question, the participant had to respond. In each situation the participants could choose between four possible ways of handling the situation. The response options were presented in writing and the participants were instructed to select the one response that was representative of how they would react. The scoring method of Ployhart et al. (2004) was followed. In contrast to many SJTs that have correct answers, the SJT used here simply records the response to each item (i.e., 1 – 4) and takes the average of all items. The answer most indicative for the communication style being measured is scored '4', and the answer least indicative for the communication style is scored '1'. An overall score of the communication style is derived from the mean of the item responses which belonged to that communication style. The total duration of the SJT was about 20 minutes.

2.2.2. CSI

To measure communication styles the CSI was used (De Vries et al., 2011). The CSI consists of 96 communication behavior items. The items are divided equally among the following six domain-level scales (16 items per scale): Expressiveness, Preciseness, Verbal Aggressiveness, Questioningness, Emotionality, and Impression Manipulativeness. Each of the domain-level scales consists of four facets, each with four items. All items were answered on a Likert-type scale with answering categories ranging from 1 (*completely disagree*) to 5 (*completely agree*). Alpha coefficients of the CSI domain-level scales ranged from .79 to .87.

3. Results Study 1

3.1. Preliminary analysis

The alpha coefficient for the SJT was .25. Before the hypothesis was tested, the correlations between the demographic characteristics and all study variables were examined. Table 1 shows the means, standard deviations, reliability coefficients, and correlations between the variables included in this study. Gender was significantly and positively related to CSI Emotionality ($r = .17, p < .05$), and significantly and negatively related to CSI Discussion Willingness ($r = -.21, p < .05$). Female participants ($M = 2.84, SD = 0.51$) showed more emotionality than male participants ($M = 3.03, SD =$

0.54, $t = 56.22$, $p < .05$). Male participants ($M = 3.27$, $SD = 0.43$) showed more discussion willingness than female participants ($M = 3.08$, $SD = 0.46$, $t = 65.99$, $p < .05$). Education was significantly and negatively related to SJT Verbal Aggressiveness ($r = -.20$, $p < .05$). Participants with level 3 of Intermediate Vocational education showed more verbal aggressiveness ($M = 1.99$, $SD = 0.42$) than participants with level 4 of Intermediate Vocational education ($M = 1.81$, $SD = 0.41$, $t = 44.34$, $p < .05$). Other demographic characteristics showed no significant correlations with the study variables.

3.2. Hypothesis testing

To test Hypothesis 1, which stated that scores on the domain-level scales of the SJT have a positive relationship with the corresponding domain-level scales of the CSI, a correlation analyses was conducted. There were no significant correlations between the scales of the SJT and the corresponding scales of the CSI, see Table 1. SJT Expressiveness was significantly and negative related to CSI Precision ($r = -.18$, $p < .05$). SJT Discussion willingness was significantly and positive related to CSI Precision ($r = .31$, $p < .01$).

Besides the correlation analysis, the factor structure of the 24 SJT items was examined. Prior to performing a principal components analysis (PCA), the suitability of data for factor analysis was assessed. First, there were no items that correlated at least .30 with one other item, suggesting a weak factorability. Secondly, the Kaiser-Meyer-Olkin measure of sampling adequacy was .54, below the recommended value of .60, and Bartlett's Test of Sphericity (Bartlett, 1954) was significant ($\chi^2 [147] = 374.12$, $p < .01$). The PCA revealed the presence of nine components with eigenvalues exceeding 1, explaining 10.1%, 7.6%, 6.9%, 6.5%, 5.6%, 5.4%, 5.1%, 4.9%, and 4.6% of the variance respectively. In the screeplot there was no clear change in the shape of the plot found. This indicates that there are no components that explain or capture much more of the variance than the other components. So, there was no clear factor structure recognized. These results do not support the use of separate SJT scales.

Table 1
Means, Standard Deviations, Scale Reliabilities, and Correlations Between Study Variables (Study 1).

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Gender	1.67	0.47	(-)																
2. Age	3.99	0.98	-.11	(-)															
3. Education	6.69	0.47	-.00	-.04	(-)														
<u>SJT</u>																			
4. Expressiveness	2.58	0.37	.04	.03	.01	(-.46)													
5. Precision	3.28	0.46	.00	.13	.04	.06	(.28)												
6. Emotionality	1.96	0.46	.15	-.08	.02	.14	.04	(.08)											
7. Discussion Willingness	2.54	0.53	-.14	.09	.04	-.08	.06	.11	(.21)										
8. Verbal Aggressiveness	1.87	0.42	-.12	-.10	-.20*	.14	-.01	-.06	-.02	(.24)									
9. Impression Manipulation	2.64	0.48	-.09	-.07	.12	.06	.07	.07	.29**	-.05	(.23)								
10. Overall SJT score	2.47	0.21	-.09	.00	.00	.41**	.41**	.39**	.56**	.42**	.52**	(.29)							
<u>CSI</u>																			
11. Expressiveness	3.15	0.46	.10	.01	.02	-.04	.06	.12	-.04	-.07	-.10	-.04	(.84)						
12. Precision	3.22	0.43	-.12	.04	.02	-.18*	-.02	.03	.31**	.04	.04	.11	-.00	(.85)					
13. Emotionality	2.97	0.53	.17*	.08	-.05	-.07	-.05	-.07	-.08	-.00	.04	-.08	-.12	-.27**	(.87)				
14. Discussion Willingness	3.12	0.46	-.21*	-.02	-.05	-.10	.08	.02	.11	.07	.07	.11	.45**	-.18*	-.13	(.84)			
15. Verbal Aggressiveness	2.59	0.44	-.09	-.05	-.12	.04	.15	-.12	-.12	.06	-.06	-.02	.13	-.29**	.23**	.19*	(.81)		
16. Impression Manipulation	2.80	0.43	-.08	.01	-.09	.06	-.01	.12	.00	.10	.08	.12	-.11	.03	.21**	.09	.08	(.79)	
17. Overall CSI score	2.98	0.21	-.07	.03	-.10	-.11	.07	.03	.06	.07	.02	.07	.49**	.21*	.40**	.64**	.49**	.47**	(.83)

Note. Scale reliabilities are presented on the diagonal, between parentheses. SJT domain-level scales were measured on a 4-point scale, and CSI domain-level scales were measured on a 5-point scale. Gender (1 = male, 2 = female), Age (2 = 18 to 25, 3 = 26 to 35, 4 = 36 to 45, 5 = 46 to 55) and Education (6 = Intermediate Vocational level 3, 7 = Intermediate Vocational level 4) were coded. $N = 147$.

* $p < .05$, ** $p < .01$

4. Discussion Study 1

The first study was meant to examine the construct validity of the video-based SJT. We hypothesized that the domain-level scales of the SJT would have a positive relationship with the domain-level scales of the CSI. This hypothesis was not supported by the results. There were no significant correlations found between the SJT scales and the corresponding domain-level scales of the CSI. However, SJT Expressiveness and SJT Discussion willingness were both significantly related to a non-corresponding domain-level scale of the CSI, namely Precision. This finding contradicts with our hypothesis, because the domain-level scales of the SJT should not significantly correlate with dissimilar domain-level scales of the CSI. Factor analysis was employed to assess if there were underlying constructs measured by the SJT. The results did not support the expectation that six constructs were measured by the SJT. The reliabilities of the SJT domain-level scales were low. However, high alpha coefficients do not necessarily imply high construct validity. In addition, alpha coefficients are usually relatively low unless the SJT is comprised of a very large number of items (Weekly & Ployhart, 2006).

Although the different steps of Ployhart et al. (2004) were precisely followed in developing the SJT, it seems reasonable to conclude that the scores of the SJT were not indicative of the constructs which were intended to measure. One possible explanation for this result could be the context of the SJT. The call center context of the SJT included customers who asked questions and expect good service. Participants probably respond in a way that they meet customer needs and have a tendency to give socially desirable answers to satisfy the customer. In addition, in most call centers there are behavioral scripts that are designed to guide employees in their interactions with customers. The non call center employees are probably also familiar with these scripts, because most people have ever been in contact with a call center employee. Therefore, it could be that participants selected response options which were most effective, instead of selecting the response options that reflect how they actually behave in a situation. Consequently it is possible that we did not measure communication styles but behavioral effectiveness.

Even though we are missing evidence for the SJT's construct validity, it is important to look at its predictive validity. It was argued earlier, that the selection of talented employees is critical to call

center success (Nicholls et al., 2009). Therefore, a second study is needed to examine if the SJT can predict job performance in a call center. Based on Study 1, it was decided to use an overall SJT score of communication instead of using the six domain-level scores.

5. Study 2

In Study 1 the construct validity of the video-based SJT was examined. The results showed that there was no evidence for the six constructs the SJT intended to measure. Subsequently, the aim of Study 2 is to examine the predictive validity of the video-based SJT. This means that it will be examined whether the SJT for communication styles is able to predict job performance of call center employees. As explained above, a total score on the SJT will be used in the subsequent analyses.

The literature suggests that the quality of communication is the most important aspect of call center work (Callaghan & Thompson, 2002). However, researchers paid minimal attention to examine the relationship between communication styles and job performance in call centers. There is evidence that communication style has a relationship with job performance, but in a leadership context. Several authors have argued that communication is a core activity of a leader (Judge, Bono, Ilies, & Gerhardt, 2002; Zaccaro, 2007). De Vries et al. (2013) provided empirical evidence for the relationship between the communication styles of the CSI and the performance of a leader. Several authors have argued that communication is a core activity for call center employees as well (Callaghan & Thompson, 2002; Webster & Sundaram, 2009), but evidence for the relationship between communication and job performance of call center employees is still missing.

Because the primary goal of call center organizations is the achievement of high levels of customers satisfaction (Jaiswal, 2008), customer satisfaction will be used as one of the job performance criteria. The other criteria are quality of work, speed of work, and the degree of following the work schedule. In general, the literature has shown that SJTs have good predictive validities. For instance, a meta-analysis of SJTs (McDaniel et al., 2001) has shown SJTs to have good predictive validity (corrected $r = .34$; $n = 10,640$). A more recent study of Lievens and Sackett (2006) showed that video-based SJTs have higher predictive validity than paper-and-pencil SJTs.

Based on these findings we hypothesized the following:

Hypothesis 2: There is a positive relationship between the overall communication score of the SJT and the job performance of call center employees.

It is also important that the video-based SJT has an added value in relation to other traditional and frequently used predictors in employee selection (Chan & Schmitt, 2002). Therefore, the third and last aim of this study is to examine the incremental validity of the SJT over cognitive ability and personality. McDaniel et al. (2007) already found in his meta-analysis that SJTs have incremental validity over cognitive ability, personality, and over a composite of cognitive ability and personality. Given the findings of McDaniel et al. and based on the literature that shows the relevance of communication for call center employees, we expect that the SJT is able to explain unique variance in job performance not explained by other measures in the test battery. Therefore the third hypothesis is:

Hypothesis 3: The SJT incrementally predicts job performance in call centers beyond personality and cognitive ability.

6. Method Study 2

6.1. Participants and Procedure

The sample consisted of 146 call center employees working in an inbound call center in the Netherlands. To increase the diversity of the sample, we asked the team leaders at the call center to select the participants on the basis of three criteria; (1) variation in level of performance, (2) variation in duration of employment, (3) variation in type of project (complex vs. less complex). Their main task was providing information in response to customer calls. The participants were working on various projects for customers of the call center (e.g., energy company, lifestyle & fashion company). Their age varied between 18 and 55 ($M = 29.34$, $SD = 9.07$). Ninety-nine participants were female (67.8%), 45 participants were male (30.8%) and two participants did not report their gender. Ninety-three participants (64.6%) had less than 12 months of work experience, 37 participants (25.7%) had more than 12 months of work experience. The selected participants received a letter with a link to the online test. The call center employees completed the tests during their working hours using their PC at the call center office. The entire predictor battery took approximately two hours to complete. First they

completed the 24-item SJT, second the 144-item measure of personality, and finally the 30 to 45-items measure of cognitive ability. The number of cognitive ability items differs because it is an adaptive test. This means that there is no fixed set of items and the next item is based on the answer on the item before. The cognitive ability measure has an item bank of several hundred items for each subtest. Because there were a lot of items in the item bank, the test would not stop if there is no stopping rule. This stopping rule is defined as follows. First, the reliability of the estimation of the test score is examined. The standard error (SE) has to be less than or equal to 0.54. Furthermore, a minimum of 10 items and a maximum of 15 items will be presented per subtest. Based on these criteria, there is a minimum of 30 questions (3 x 10) and maximum of 45 questions (3 x 15).

Job performance measures of the call center employees were obtained via their team leaders. In total, 18 team leaders from the call center evaluated the call center employees on four job performance criteria. SJT scores of the non call center employees from Study 1 were also included in the dataset to be able to compare groups.

6.2 Measures

6.2.1 SJT

The same SJT as described in Study 1 was used but in this study a different scoring method was used, namely an expert-based scoring method. This method was used to determine how well the test takers did on the SJT. In addition, the results of Study 1 did not support the use of a construct-based scoring method. SJT scores of above average performing call center employees were used as expert-ratings. Scores on each item could range from 0 to 4 and were calculated as follows. The response option that the majority of the experts had chosen received 4 points. Next, the percentages of others responses were calculated and five scoring categories were created. Suppose the most frequently chosen answer for an item is chosen by 50% of the employees, then the following categories were created. A '0' was received if 0-10% of the employees chose the response option, a '1' for 10-20%, a '2' for 20-30% and a '3' was rated when 30-40% of the employees chose that response option. This scoring key was applied to each single item. To determine the SJT scores of the non call center employees, they were compared with the SJT scores of the high performing call center employees. A total test score was created by summing up the scores on all items.

6.2.2. Reflector Big Five Personality (RBFP)

We used the RBFP (Schakel, Smid, & Jaganjac, 2007) to measure personality. The RBFP is an online computer-based Big Five personality questionnaire applied to situations and behavior in the workplace. It is a Dutch version of the Workplace Big Five Profile constructed by Howard and Howard (2001). This profile is based on the NEO-PI-R (Costa & McCrae, 1992) and adapted to workplace situations. It consists of 144 items, distributed over five scales (Need for Stability, Extraversion, Openness, Agreeableness, and Conscientiousness). The items are scored on a five point Likert scale (1 = *least indicative for the trait*, 5 = *most indicative for the trait*). Coefficient alphas varied from .65 to .88 for the five scales.

6.2.3. Cognitive Ability Test

Cognitive ability was assessed with the Connector Ability (Maij- de Meij, Schakel, Smid, Verstappen, & Jaganjac, 2008). The Connector Ability measures general cognitive ability level by means of three subtests: Figure Series (FS), Matrices (M), and Number Series (NS). The Connector Ability aims at Intermediate Vocational education. A minimum of 10 items and a maximum of 15 are presented per subtest. The reliabilities (coefficient alphas) of the subtests were .78 for the FR scale, .73 for the M scale, .88 for the NS scale.

6.2.4. Job Performance

Team leaders provided performance ratings of 130 participants. Performance indicators for this call center included customer satisfaction, quality of work, speed of work, and the degree of following work schedule. Each domain was measured with one item and rated on a scale from 1 (*low*) to 5 (*high*). An overall measure of job performance was created by averaging the four ratings. On the basis of standardized total scores, the group was divided into below average and above average performing employees. Below average were the standardized scores till 0, above average was a z-score of 0 or higher.

7. Results Study 2

7.1. Preliminary analysis

Before testing the hypotheses, the correlations between the demographic characteristics and all study variables were examined. Table 2 presents the means, standard deviations, reliability coefficients,

and correlations between the variables included in this study. Gender was significantly and positively related to the SJT ($r = .19, p < .05$), to agreeableness ($r = .32, p < .01$), and to conscientiousness ($r = .19, p < .05$). Female participants ($M = 80.03, SD = 5.46$) scored significantly higher on the SJT than male participants ($M = 77.64, SD = 6.26; t = -2.31$). Female participants ($M = 46.84, SD = 8.53$) showed significantly more agreeableness than male participants ($M = 39.46, SD = 13.49; t = -3.84$). Female participants ($M = 52.57, SD = 8.90$) also showed significantly more conscientiousness than male participants ($M = 48.58, SD = 10.98, t = -2.23$). Age was significantly and positively related to agreeableness ($r = .18, p < .05$) and to conscientiousness ($r = .19, p < .05$). Duration of employment had a positive and significant correlation with customer satisfaction ($r = .35, p < .01$), quality of work ($r = .38, p < .01$), degree of following work schedule ($r = .32, p < .01$), and overall job performance ($r = .36, p < .01$). Complexity of project was significantly and negatively related to extraversion ($r = -.20, p < .05$). Because of these significant correlations, we controlled for gender, age, duration of employment and complexity of project in the regression analyses.

7.2. Hypotheses testing

Hypothesis 2 stated that the overall communication score of the SJT has a positive relationship with job performance of call center employees. To test this hypothesis a correlation analysis was conducted. Scores on the SJT showed no significant correlations with the overall job performance or with one of the single job performance measures, see Table 2. These findings did not support our second hypothesis. To examine if there was a significant difference between the SJT scores for call center employees and non call center employees, an independent-samples *t*-test was conducted. Call center employees scored significantly higher on the SJT ($M = 78.95, SD = 6.14$) than non call center employees ($M = 71.34, SD = 8.18; t = 8.55, p < .01$). Above average performing call center employees ($M = 79.75, SD = 6.25$) scored significant higher than non call center employees ($M = 71.33, SD = 8.18; t = 7.49, p < .01$). Under average performing call center employees ($M = 78.02, SD = 5.93$) scored also significant higher than non call center employees ($M = 71.33, SD = 8.12; t = 5.63, p < .01$). Above average performing call center employees ($M = 79.75, SD = 6.25$) did not score significantly different than under average performing call center employees ($M = 78.02, SD = 5.93; t = -1.57, p$

= .12). These results indicate that the SJT has the potential to distinguish between experienced call center employees and non experienced call center employees.

The third and last hypothesis stated that the SJT incrementally predicts job performance in call centers beyond personality and cognitive ability. Hierarchical regression analyses were conducted to test this hypothesis. Step 1 included gender, age, duration of employment, and complexity of project. Step 2 included the Big Five personality dimensions, Step 3 included cognitive ability, and the final step included the SJT. For each job performance measure, the same regressions were used. The results are presented in Table 3. The SJT was not able to explain significant variance in any of the job performance measures. Regarding customer satisfaction, significant beta weights were found for cognitive ability ($\beta = -.18, p < .05$) and duration of employment ($\beta = .39, p < .05$). Regarding quality of work, speed of work, degree of following work schedule and overall job performance, significant beta weights were found for duration of employment ($\beta = .43, p < .01, \beta = .19, p < .05, \beta = .33, p < .01$ and $\beta = .39, p < .05$ respectively). Personality and cognitive ability were not able to explain significant variance in the job performance measures. All predictors explained 18% of the variance in customer satisfaction ($F = 2.25, p < .05$), 21% of the variance in quality of work ($F = 2.81, p < .01$), 21% of the variance in degree of following work schedule ($F = 2.17, p < .05$), and 17% of the variance in overall job performance ($F = 2.14, p < .05$). The predictors were not able to explain significantly variance in speed of work ($F = 0.19, p = .54$). Based on these findings, Hypothesis 3 could not be supported.

Table 2
Means, Standard Deviations, Scale Reliabilities, and Correlations Between Study Variables (Study 2).

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Gender	1.69	0.47	(-)															
2. Age	29.34	9.07	-.08	(-)														
3. Duration of employment	1.28	0.45	-.16	.27**	(-)													
4. Complexity of project	1.86	0.54	-.11	.13	.18*	(-)												
<u>Predictors</u>																		
5. SJT	79.06	6.19	.19*	.03	.12	.03	(-)											
6. Need for stability	51.13	7.76	.10	-.03	.08	.16	-.02	(.81)										
7. Extraversion	46.79	8.12	.10	-.01	-.15	-.20*	-.07	-.33**	(.73)									
8. Openness	44.96	8.42	-.09	.11	-.02	-.01	-.07	-.41**	.48**	(.74)								
9. Agreeableness	44.72	10.77	.32**	.18*	.11	-.06	.05	-.11	.30**	.24**	(.65)							
10. Conscientiousness	51.59	9.83	.19*	.19*	-.06	-.05	-.00	-.28**	.42**	.69**	.35**	(.88)						
11. Cognitive Ability	46.97	7.64	.05	.15	-.09	-.09	.32**	.10	-.07	.02	.03	-.00	(-)					
<u>Job performance measures</u>																		
12. Customer satisfaction	0.00	0.98	.00	-.01	.35**	.02	-.02	.04	.03	.04	.06	-.03	-.21*	(-)				
13. Quality of work	0.00	0.98	.06	.03	.38**	.01	.03	-.03	.05	.09	.05	-.03	-.10	.68**	(-)			
14. Speed of work	0.00	0.98	-.10	-.07	.17	.01	-.05	.01	.07	.10	.07	.03	-.09	.52**	.54**	(-)		
15. Degree of following work schedule	0.00	0.98	.08	.13	.32**	.04	.17	-.06	-.12	-.03	.07	-.13	-.10	.67**	.51**	.46**	(-)	
16. Overall job performance	0.00	0.98	-.03	.00	.36**	.01	.02	-.03	.05	.11	.05	.01	-.14	.79**	.86**	.82**	.71**	(.88)

Note. Scale reliabilities are presented on the diagonal, between parentheses. Gender (1 = male, 2 = female), duration of employment (1 = less than one year, 2 = more than one year), and complexity of project (1 = less complex, 2 = average complex, 3 = more complex) were coded. Personality scales were measured on a 5-point scale. Scores on the SJT had a maximum of 96, and job performance measures were measured on a 4-point scale and were then standardized (*z*-scores). *N* = 144 for gender and SJT. *N* = 147 for age. *N* = 132 for duration of employment, complexity of project, quality of work and speed of work. *N* = 137 for Big Five personality scales. *N* = 139 for cognitive ability. *N* = 126 for customer satisfaction and overall job performance, and *N* = 102 for degree of following work schedule.

* $p < .05$, ** $p < .01$

Table 3
Hierarchical Regression Analyses of Predictors on the Job Performance Measures (Study 2).

	Customer satisfaction		Quality of work		Speed of work		Degree of following work schedule		Overall job performance	
	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2
<u>Step 1</u>										
Gender	.05		.11		-.08		.13		.02	
Age	-.11		-.07		-.13		.05		-.10	
Duration of employment	.39**		.43**		.19*		.33**		.39**	
Complexity of project	-.04	.14	-.05	.17	-.01	.05	-.01	.12	-.04	.14
<u>Step 2</u>										
Extraversion	.08		.09		.07		-.11		.06	
Agreeableness	-.07		-.07		.08		.02		-.03	
Conscientiousness	-.68		-.20		-.01		-.29		-.11	
Emotional stability	.04		-.03		.07		-.15		-.00	
Openness to experience	.10	.01	.21	.04	.10	.03	.16	.06	.19	.03
<u>Step 3</u>										
Cognitive ability	-.18*	.03	-.07	.00	.06	.00	-.11	.01	-.11	.01
<u>Step 4</u>										
SJT	.00	.00	-.01	.00	-.02	.00	.16	.02	.03	.00
R^2		.18		.21		.08		.21		.17
F		2.25*		2.81**		0.91		2.17*		2.14*

Note: Gender (1 = male, 2 = female), duration of employment (1 = less than one year, 2 = more than one year), and complexity of project (1 = less complex, 2 = average complex, 3 = more complex) were coded. $N = 144$ for gender and SJT. $N = 147$ for age. $N = 132$ for duration of employment, complexity of project, quality of work and for speed of work. $N = 137$ for Big Five personality scales. $N = 139$ for cognitive ability. $N = 126$ for customer satisfaction and overall job performance, and $N = 102$ for degree of following work schedule.

* $p < .05$, ** $p < .01$

8. General discussion

The aim of the present study was to develop a video-based SJT for the selection of call center employees and to investigate its construct validity, predictive validity, and incremental validity. Study 1 involved the development of a construct-driven video-based SJT. The SJT was developed on the basis of an existing communication style model, namely the CSI (De Vries et al., 2011). To examine if the SJT actually measured the constructs it intended to measure, its construct validity was investigated. This was done by examining the relationship between the scores on the domain-level scales of SJT and the scores of the corresponding domain-level scales of the CSI. There were no significantly relationships found between the scores of the corresponding domain-level scales. Furthermore, the results of the factor analysis did not confirm the presence of six communication styles. Consequently the first hypothesis was not supported.

The results can possibly be explained by the different characteristics of the SJT and the CSI. First, the response instructions of the SJT and the CSI differ. In the CSI questionnaire, participants were instructed to indicate to what extent they agree with a statement. In the SJT, participants were instructed to identify what he or she would do given the situation, what could be labeled as a 'behavioral tendency' (would do) instruction type. Response instructions are likely to affect construct validity (Weekley & Ployhart, 2006). McDaniel and Nguyen (2001) argued that SJTs with behavioral tendency instructions are more susceptible to faking because applicants may be motivated to select response options that are socially desirable, even if the options do not correspond to what they would typically do at work. The response instruction of the CSI is less likely to evoke socially desirable behavior. In addition, the call center context of the SJT could probably elicit more socially desirable behavior than the context-free statements in the CSI. Together, the context and the response instructions of the SJT could lead to developing beliefs about which behavior would be most effective in a given situation. Therefore we probably measured behavioral effectiveness instead of measuring communication styles. Overall, these findings suggest that SJTs could be more appropriate to measure constructs like personality or job knowledge, than measuring other constructs like communication.

In the first study, the development procedure of Ployhart et al. (2004) was followed. They found evidence for the construct validity of the SJT, which was an important reason why we used

their development procedure in this study. This procedure included a scoring method that was in contrast to many SJTs that have correct answers. The response options in this SJT reflected a range of the specific communication style within a given situation. The answer most indicative for the communication style was scored '4' and the answer least indicative for the communication style was scored '1'. Although the SJT was carefully constructed and followed all the procedural steps, there was no evidence found for the construct validity of the SJT. Therefore it was decided to use an expert-based scoring method in the second part of the study. Scores of above average performing call center employees were used to create an overall SJT score which was used in the analyses.

Study 2 involved the investigation of the predictive validity of the SJT. This was done by examining the relationship between the overall SJT score and several job performance measures of call center employees. The results showed that the overall SJT score was not related to any of the job performance measures. Therefore, no support was found for the second hypothesis. The results were surprising because several studies have shown that SJTs have good predictive validities (e.g., McDaniel et al., 2001; Lievens & Sackett, 2006). Our findings could possibly be explained by the way the job performance ratings were done. A total of eighteen team leaders rated the job performance of the call center employees. Some team leaders judged just one employee, but other team leaders judged a lot more employees. Team leaders did not rate the performance of their personnel with an objective performance measurement system, so each team leader used their own subjective norm for the job performance ratings. Subjectivity in performance evaluations has some potential disadvantages. Raters have a natural self-serving bias, a tendency to inflate their subordinate's ratings so that they appear to be successful (Greenberg, 1991). In addition, most evaluators prefer to have a pleasant relationship with their subordinates, which is sometimes referred to as a desire to minimize confrontation costs (Bol, 2008; Varma, DeNisi, & Peters, 1996). Furthermore, the call center employees worked on seven different projects which differed in complexity. Therefore it was difficult to compare the job performance ratings which each other. Overall, the way the performances of the employees were rated, could have influenced the results regarding the predictive validity of the SJT.

Based on the literature, it was logical to expect that more variables were positively related to the job performance measures. For example, several studies found that conscientiousness is a valid

predictor of job performance in all occupational groups (e.g., Mount & Barrick, 1998; Avis, Kudisch, & Fortunato, 2002). In this study, there were no significant relationships found between conscientiousness and the job performance measures. Beside this, cognitive ability was also not significant related to job performance. Regarding customer satisfaction, a significant en negative beta value was found for cognitive ability which means that higher cognitive ability of employees would lead to less customer satisfaction. Given the overwhelming research that shows the strong link between cognitive ability and job performance (Hunter, 1986), this result was not expected. These findings could probably also be explained by the expert ratings. Duration of employment was the only variable that was significantly and positively related to job performance. This was in line with previous findings of McDaniel, Schmidt, and Hunter (1988) who found a relationship between job experience and job performance. In their study, job experience was defined as length of experience in a given occupation.

In Study 2 the difference between the overall SJT score of the call center employees and non call center employees was also examined. Call center employees scored significantly higher on the SJT than non call center employees. This significant difference could possibly be explained by previous findings that SJTs are measures of job knowledge (Clevenger et al., 2001, Weekley & Ployhart, 2004). Obviously call center employees have more call center specific job knowledge than non call center employees (MacKenzie, Ployhart, Weekley, & Ehlers, 2010). The gender differences found at the SJT level are consistent with prior research on SJTs, namely that woman typically score higher than men (Motowidlo et al., 1990; Motowidlo & Tippens, 1993; Weekly & Jones, 1999). Weekly and Jones (1999) argued that the interpersonal nature of many problems in SJTs tend to favor woman.

Finally, the incremental validity of the SJT was investigated. It was hypothesized that the SJT would incrementally predict job performance beyond personality and cognitive ability. The results were in contrast with our hypothesis. The results showed that the SJT did not explain unique variance in the job performance scores beyond personality and cognitive ability. These results are not in line with previous research of McDaniel et al. (2007) who found in their meta-analysis that SJTs have incremental validity over cognitive ability and personality. An explanation for this result is that the overall SJT score was not significantly related to the job performance measures. Besides that, the SJT

showed a significant and positive relationship with cognitive ability, which can explain why the SJT had no incremental validity over cognitive ability. Cognitive ability was also not able to explain unique variance in job performance. It was mentioned earlier that this result is quite surprising because of prior research findings that shows the positive relation between cognitive ability and job performance (Hunter & Hunter, 1984; Hunter & Schmidt, 1996). In addition, several studies found strong relationships between personality traits and job performance (Mount & Barrick, 1998; Barrick & Mount, 1991), but in this study personality did not explain unique variance in job performance. It could be that personality questionnaires may be more predictive for certain job categories than for others, but the fact that none of the traditional measures were related to job performance provides further evidence that the problems with the job performance measures have attenuated the validity results.

The present study is one of the first that developed a construct-based SJT for communication styles. Secondly, there are no prior studies that developed a SJT for the selection of call center employees. Therefore, we believe that this study makes a contribution to the literature. Overall, the results did not provide support for the construct validity of the SJT, and we suggested that SJTs are probably not appropriate to measure a construct like communication style. Furthermore, there was no evidence found for the predictive validity of the SJT. It was suggested that this is probably due to the job performance measures. The study demonstrates that call center employees scored higher on the SJT than non call center employees and it has the potential to distinguish between experienced and non experienced people. When using the SJT in a selection procedure, the applicants who will not meet the minimum score of a call center employee will probably not fit in a call center.

Limitations and directions for future research

There are some potential limitations in this study that must be considered. First, in both studies we used participants from an Internet panel which were not real applicants for a call center job. Considerable research suggests that applicants are more motivating than anonymous/voluntary participants and may lead to differences in socially desirable responding (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Ployhart, Weekley, Holtz, & Kemp, 2003). Using participants who were not working in a call center limits the extent to which the findings can be generalized to an

actual job applicant sample. Therefore, it is recommended to use actual applicants in future research.

Second, the results of the call center participants represented just one call center. Further research should include other call centers to generalize the findings. Third, there were missing data in the performance outcomes. Some team leaders rated their employees on just one job performance outcome, and other team leaders did not rate their employees. This reduced the sample size for the analysis of validity against job outcomes. It was mentioned earlier that the job performance ratings probably were not valid because of the subjectivity of the ratings. Another limitation is that the CSI was not included in Study 2. So, we do not have CSI scores of the call center employees and consequently do not know if this questionnaire is able to predict job performance. Further research should attempt to address these limitations.

Based on the results of the Study 1 we believe that is difficult to develop a construct-based SJT for communication styles. The steps of the development process were carried out very precisely, but we did not find evidence that the SJT measured communication styles. However, we still believe that someone's style of communication is a key factor for customer satisfaction, because it connects employees and customers (Mohr & Navin, 1990). Achieving greater customer satisfaction is important for an organization to distinguish themselves from all the other call centers. Therefore, we recommend future studies to investigate the influence of communication on job performance, but using other communication measures. In Study 2 it was demonstrated that scores of experienced call center employees were significantly higher than non call center employees. Further research is necessary to investigate if this difference is also present when non call center employees are replaced by real applicants. Moreover, the SJT can be used in a selection procedure, because with the SJT we are able to measure if an applicant is likely to behave the same as a call center employee. When the SJT has been used for a while in a selection procedure, it would be recommended to investigate the job performance of the selected applicants. In this way, it would be clear if the SJT is really able to select applicants who perform well as a call center employee.

References

- Aksin, O. Z., & Harker, P. T. (1999). To sell or not to sell: Determining trade-offs between services and sales in retail banking phone centers. *Journal of Service Research, 2*, 19-33.
- Arthur, W. Jr., & Villado A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection and research in practice. *Journal of Applied Psychology, 93*, 435–442.
- Ashton, M. C., & Lee, K. (2008). The prediction of honesty-humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality, 42*, 1216-1228.
- Avis, J. M., Kudisch, J. D., & Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology, 17*, 87-105.
- Bain, P., & Taylor, P. (2000). Entrapped by the ‘electronic panopticon? Worker resistance in the call centre. *New Technology, Work, and Employment, 15*, 2-18.
- Bakker-Pieper, A., & De Vries, R. E. (2013). The incremental validity of communication styles over personality traits for leader outcomes. *Human Performance, 26*, 1-19
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Bartlett, M.S. (1954). A note on the multiplying factors for various chi-square approximations. *Journal of the Royal Statistical Society, 16*, 296–298.
- Bauer, T., & Truxillo, D. (2006). Applicant reactions to situational judgment tests: Research and related practical issues. In J. A. Weekley & R. E. Ployhart (Eds.). *Situational judgment tests: Theory, measurement and application* (pp. 233–247).
- Bloomquist, M. J., & Kleiner, B. H. (2000). How to reduce theft and turnover through better hiring methods. *Management Research News, 23*, 79-83.
- Bodin, M., & Dawson, K. (1999). *The call centre dictionary*. New York: Telecom Books.
- Bol, J. C. (2008). Subjectivity in compensation contracting. *Journal of Accounting Literature, 27*, 1-32.
- Callaghan, G., & Thompson, P. (2002). We recruit attitude: The selection and shaping of routine call centre labour. *Journal of Management Studies, 39*, 234- 254.

- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*, 233-254.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83-117.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt-Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410-417.
- Costa, P. T., & McCrae, R. R. (1992). *NEO Personality Inventory—Revised (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- De Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO persoonlijkheidsvragenlijst [The six most important personality dimensions and the HEXACO Personality Inventory]. *Gedrag & Organisatie, 22*, 232-274.
- De Vries, R. E., Bakker-Pieper, A., Konings, F. E., & Schouten, B. (2011). The Communication Style Inventory (CSI): A six-dimensional behavioral model of communication styles and its relation with personality. *Communication Research, 40*, 506-532.
- De Vries, R. E., Bakker-Pieper, A., Alting Siberg, R., Van Gameren, K., & Vlug, M. (2009). The content and dimensionality of communication styles. *Communication Research, 36*, 178-206.
- Greenberg, J. (1991). Motivation to inflate performance ratings: Perceptual bias or response bias? *Motivation and Emotion, 15*, 81-97.
- Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *Handleiding bij de NEO persoonlijkheidsvragenlijsten NEO-PI-R en NEO-FFI* (Manual of the NEO personality inventories NEO-PI-R and NEO-FFI). Lisse, Netherlands: Swets & Zeitlinger.

- Hough, L. M., Eaton N. K., Dunnette M. D., Kamp J. D., & McCloy R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.
- Howard, P.J., & Howard, M.J. (2001). *Professional manual for the Workplace Big Five profile (WB5P)*. Charlotte, NC: Centacs.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.
- Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2*, 447-472.
- Hutchinson, S., Purcell, J., & Kinnie. N. (2000). The challenge of the call center. *Human Resource Management International Digest, 8*, 4-8.
- Jaiswal, A. K. (2008). Customer satisfaction and service quality measurement in Indian call centres. *Managing Service Quality, 18*, 405-416.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765–780.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39*, 329-358.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*, 426–441.
- Lievens, F., Van Dam, K., & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review, 31*, 580–601.
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181–1188.
- Ma, J., Kim, N., & Rothrock, L. (2011). Performance assessment in an interactive call center workforce simulation. *Simulation Modelling Practice and Theory, 19*, 227-238.

- MacKenzie, W. I., Ployhart, R. E., Weekley, J. A., & Ehlers, C. (2010). Contextual effects on SJT responses: An examination of construct validity and mean differences across applicant and incumbent contexts. *Human Performance, 23*, 1-21.
- Maij-de Meij, A. M., Schakel, L., Smid, N., Verstappen, N., & Jaganjac, A. (2008). *Connector Ability 1.1, Professional Manual*. Utrecht: PiCompany B.V.
- Mattila, A. S., & Mount, D. J. (2003). The role of call centers in mollifying disgruntled guests. *Cornell Hotel and Restaurant Administration Quarterly, 44*, 142-50.
- McCroskey, J. C., & Beatty, M. J. (2000). The communibiological perspective: Implications for communication in instruction. *Communication Education, 49*, 1-6.
- McCulloch, M. C., & Turban, D. B. (2007). Using Person-Organization Fit to select employees for high-turnover jobs. *International Journal of Selection and Assessment, 15*, 63-71.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.
- 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 test: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology, 73*, 327-330.
- Messick S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Mohr, J., & Nevin, J. R. (1990). Communication strategies in marketing channels: A theoretical perspective. *Journal of Marketing, 54*, 36-51.
- Morris, S. B., & Lobsenz, R. E. (2002). Significance tests and confidence intervals for the adverse impact ratio. *Personnel Psychology, 53*, 89-111.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

- Motowidlo, S. J., & Tippens, N. (1993). Future studies of the low-fidelity stimulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, *66*, 337-344.
- Mount, M. K., & Barrick, M. R. (1998). Five reasons why the 'Big Five' article has been frequently cited. *Personnel Psychology*, *51*, 849-857.
- Nicholls, M., Viviers, A. M., & Visser, D. (2009). Validation of a test battery for the selection of call centre operator in a communication company. *South African Journal of Psychology*, *39*, 19-31
- Peccei, R., & Rosenthal, P. (1997). The antecedents of employee commitment to customer service: Evidence from a UK service context. *International Journal of Human Resource Management*, *8*, 66-86.
- Philips, J. M. (1998). Effects of realistic job previews on multiple organizational outcomes: A meta-analysis. *Academy of Management Journal*, *41*, 673-690.
- 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 & Assessment*, *11*, 1-16.
- Ployhart, R. E., Porr, W. B., & Ryan, A. M. (2004). A construct-oriented approach for developing situational judgment tests in a service context. Unpublished manuscript.
- Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: are personality, biodata, and situational judgment tests comparable? *Personnel Psychology*, *56*, 733-752.
- Prabhaker, P. R., Sheehan, M. J., & Coppett, J. I. (1997). The power of technology in business selling: call centers. *Journal of Business & Industrial Marketing*, *12*, 222-235.
- Ramseook-Munhurrun, P., Naidoo, P., & Lukea-Bhiwajee, S. D. (2009). Employee perceptions of service quality in a call centre. *Managing Service Quality*, *19*, 541-557.
- Ring, P. S., & Van de Ven, A. H. (1994). Developmental processes of cooperative interorganizational relationships. *Academy of Management Review*, *19*, 90-118.
- Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, *85*, 880-887.

- Sawyer, O. O., Srinivas, S., & Wang, S. (2009). Call center employee personality factors and service performance. *Journal of Services Marketing, 23*, 301-317.
- Schakel, L., Smid, N. G., & Jaganjac, A. (2007). *Workplace Big Five professional manual*. Utrecht, The Netherlands: PiCompany B.V.
- Shoda, Y., Mischel, W., & Wright, J. C. (1993). The role of situational demands and cognitive competencies in behavior organization and personality coherence. *Journal of Personality and Social Psychology, 65*, 1023-1035.
- Smith, W. L. (2001). Customer service call centers: Managing rapid personnel changes. *Human Systems Management, 20*, 123-129.
- Varma, A., Denisi, A. S., & Peters, L. H. (1996). Interpersonal affect and performance appraisal: A field study. *Personnel Psychology, 49*, 341-360.
- Webster, C., & Sundaram, D. S. (2009). Effect of service provider's communication style on customer satisfaction in professional services setting: The moderating role of criticality and service nature. *Journal of Service Marketing, 23*, 104-114.
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700.
- Weekley, J. A., Ployhart, R. E., & Harold, C. M. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance, 17*, 433-461.
- Weekley, J. A., & Ployhart, R. E. (2006). *Situational Judgment Tests: Theory, measurement, and application*. Mahwah, NJ: Lawrence Erlbaum.
- Weekley, J. A., & Ployhart, R. E. (2004). Situational judgment: Antecedents and relationships with performance. *Human Performance, 18*, 81-104.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188-202.
- Zaccaro, S. J. (2007). Trait-based perspectives of leadership. *American Psychologist, 62*, 6-16.
- Zapf, D., Isic, M., Bechtoldt, M., & Blau, P. (2003). What is typical for call centre jobs? Job characteristics and service interactions in different call centres. *European Journal of Work and Organizational Psychology, 12*, 311-340.