Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

4-2023

Arousing motives or eliciting stories? On the role of pictures in a picture–story exercise

Philipp Schäpers
University of Munster

Stefan Krumm
Freie Universitat Berlin

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Nikola Stenzel
Psychologische Hochschule Berlin

Article

Arousing Motives or Eliciting Stories? On the Role of Pictures in a Picture–Story Exercise

Philipp Schäpers[1]* , **Stefan Krumm[2]***, **Filip Lievens[3]**, and **Nikola Stenzel[4]**

Abstract

Picture–story exercises (PSE) form a popular measurement approach that has been widely used for the assessment of implicit motives. However, current theorizing offers two diverging perspectives on the role of pictures in PSEs: either to elicit stories or to arouse motives. In the current study, we tested these perspectives in an experimental design. We administered a PSE either with or without pictures. Results from $N = 281$ participants revealed that the experimental manipulation had a medium to large effect for the affiliation and power motive domains, but no effect for the achievement motive domain. We conclude that the herein chosen pictures cues function differentially across motives, as they aroused the affiliation and power motives, but not the achievement motive.

Research on implicit motives has a long tradition and fascinated researchers and practitioners alike. In this context, picture–story exercises (PSEs) form one popular measurement approach that has been widely used for the assessment of implicit motives. The reasons for their popularity are manifold: For instance, PSEs seek to gauge personal information individuals might not even be aware of themselves (e.g., McClelland, 1985; Schönbrodt et al., 2020), they are less prone to socially desirable response behavior than self-report questionnaires (Gruber & Kreuzpointner, 2015), and show criterion-related validity for operant outcomes (e.g., entrepreneurial performance or career choice, Collins et al., 2004).

PSEs denote a storytelling technique that consists of ambiguous pictures which are presented to test takers (Schultheiss & Pang, 2007). Pictures typically show social daily-life situations, such as two people standing at a bar or an old woman standing behind a young woman and looking at her. Pictures are intentionally designed so that they are open to test takers' subjective interpretations. Next, individuals are asked to write a story related to each picture. These stories are then analyzed by trained coders on the basis of established coding manuals and inferences are made about test takers' achievement, power, and affiliation motives (e.g., Winter, 1994).

However, theorizing around PSEs offers different explanations on how responses are formed (Tuerlinckx et al., 2002). On one hand, many authors posit that pictures arouse motives which then manifest in the stories written by the

test takers (for an overview, see Pang, 2010). So, several authors assign a causal role to pictures in terms of eliciting motives (as also reflected in their categorization as stimulus attribution tests; Bornstein, 2011). On the other hand, on the basis of motive theories that consider motives as dynamic and constantly competing with each other (Atkinson & Birch, 1970), one can also conclude that motives are momentarily aroused or satisfied by many variables beyond pictures or can even be constantly (sometimes also called "chronically") aroused (for an overview, see Schultheiss et al., 2010). According to this view, the role of pictures primarily is to make individuals tell a story, but not to arouse motives (also see classical experiments by Atkinson & McClelland, 1948). Notably, some of the previous literature mixed both views and interchangeably attributed both roles to picture cues (arousing motives and eliciting stories) as if they were the same (e.g., Weiner & Greene, 2017). In the current study, we argue that these are opposing views

[1]University of Münster, Münster, Germany
[2]Freie Universität Berlin, Berlin, Germany
[3]Singapore Management University, Singapore
[4]Psychologische Hochschule Berlin, Berlin, Germany

*Philipp Schäpers and Stefan Krumm contributed equally to this manuscript.

Corresponding Author:
Philipp Schäpers, Department of Psychology, University of Münster, Fliednerstraße 21, Münster 48149, Germany.
Email: philipp.schaepers@googlemail.com

and examine the relevance of pictures for the motive-related content in a PSE. In doing so, our study contributes to a deeper understanding of PSEs.

Study Background

Picture–Story Exercises

As one of the most prominent PSEs, the Thematic Apperception Test (TAT; C. D. Morgan & Murray, 1935) was first introduced to the scientific community in 1935 by C. D. Morgan and Murray. Like other PSEs, it has since then been widely used in practice and in the social sciences (Childs & Eyde, 2002; Piotrowski, 2017). For instance, researchers show high interest in using PSEs for inspecting individual differences in implicit aspects of the personality (e.g., Baumann et al., 2005; Slabbinck et al., 2013) or to monitor the psychotherapy progress of a client (Weiner & Greene, 2017).

Although being the most prominent proponent, the TAT is by far not the only PSE that has been developed (e.g., see Bernecker & Job, 2011; Costantino et al., 2014; Costantino & Malgady, 2008; George & West, 2001; Runge & Lang, 2019; Schultheiss & Pang, 2007). In fact, PSE refers to an assessment technique in which ambiguous pictures are presented and test takers are asked to come up with an imaginative story. In more general terms, this technique may also be classified as picture-based projective testing (Kubiszyn et al., 2000).

Previous applications of PSEs (e.g., the TAT) mostly focused on the so-called Big-Three motives (Kehr, 2004), that is, on achievement, affiliation, and power motives. Motives denote a person's tendency to direct and sustain her or his behavior so that specific goal-states are achieved, which then results in motive satisfaction (Schultheiss, 2008). The nature of goals and incentives that result in motive satisfaction can be quite different and may, briefly described, range from the completion of a complex task (achievement motive) to sustaining personal relationships (affiliation motive) to dominating others (power motive; Smith et al., 1992).

Theorizing on motives also emphasized the distinction between a person's *conscious* versus *unconscious* preference for (achievement-, affiliation-, and power-related) goal states (McClelland et al., 1989). Conscious individual representations of motives, as assessed with self-report questionnaires, have been termed explicit motives. Conversely, unconscious preferences to direct behavior to specific (pleasant) goal states are referred to as implicit motives. For instance, individuals with a strong power motive may seek opportunities to control others and enjoy doing so without being aware of this tendency. It follows that implicit motives cannot be assessed through self-reports (Köllner & Schultheiss, 2014), but call for an indirect assessment—as provided by PSEs.

Psychometric Properties of PSEs. Mirroring the longstanding tradition of PSEs, ample research examined the psychometric properties of the scores obtained. For instance, it has long been argued that two or more human coders come to rather different conclusions regarding the motive-related content in the same written stories (for further information on the critiques of PSEs, see, e.g., R. J. Jenkins, 2017; Lilienfeld et al., 2000). However, elaborated scoring manuals were developed, which—together with extensive training of coders—allow an objective scoring of test takers' personality on the basis of written stories (e.g., see Smith et al., 1992; Winter, 1994). Much the same is true for PSEs' reliability estimates: If elaborate item response models (Lang, 2014) or the appropriate number of picture stimuli (Hibbard et al., 2001) are applied, satisfactory reliability estimates can be achieved (cf. see also Lundy, 1985, for a discussion that Cronbach's α might not be an appropriate estimate for PSEs' reliability).

Arguably, most research efforts were put into examining the correlation of PSEs with other variables, thus providing evidence for the validity of conclusions drawn from PSEs. A meta-analysis by Köllner and Schultheiss (2014) summarized research on the convergence between motives assessed with PSEs and self-report questionnaires. In line with dual motive theory (McClelland et al., 1989), Köllner and Schultheiss found that, overall, explicit and implicit assessments of the same motive show a correlation of only .11 (corrected for sampling and measurement error). Furthermore, there is a substantial body of literature that looked at criteria that were related to PSE or, more specifically, TAT scores. In an earlier meta-analysis, for instance, Spangler (1992) examined the correlation of the achievement motive, as measured with the TAT, with a variety of criteria. As hypothesized, he revealed that TAT achievement motive scores were particularly predictive of real-life criteria (such as sales success or income) when achievement-related activity incentives were present (e.g., time pressure). Further evidence is provided by the meta-analysis of Collins et al. (2004) who found that the achievement motive, measured with the TAT, was significantly correlated with interest in entrepreneurship and performance as an entrepreneur (for further evidence on the criterion-related validity, see also Bornstein, 1999).

Although research on correlations between PSEs and other variables has cumulated in various meta-analyses, research on response processes, as another source of validity (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), is relatively sparse and mostly focus on fitting item response models to responses obtained from PSEs (e.g., Blankenship et al., 2006; Gruber & Kreuzpointner, 2013; Lang, 2014; Tuerlinckx et al., 2002). However, this research neglected to provide insight into the process between item administration and item response and has not yet addressed a

fundamental question: Do pictures in PSEs *arouse motives and elicit stories* or do they *only elicit stories*? Next, we delineate both of these views in more detail.

Pictures Arousing Motives and Eliciting Stories. In line with the notion that PSEs are projective tests, a core rationale is that ambiguous pictures enable test takers to project their motives into the presented pictures and thus write or tell a story reflecting these motives (Frank, 1939; Schultheiss & Brunstein, 2001). For this test principle to work, it is essential that the chosen pictures can indeed arouse motives. The extent to which a motive is aroused in an individual—and then manifest in stories—is used as an indicator of implicit motive strength. The average extent to which picture cues elicit motive-related story content is also referred to as card pull (Peterson & Schilling, 1983; Stein et al., 2014), cue strength (Smith et al., 1992), or instigating force (Tuerlinckx et al., 2002).

From a psychometric perspective, card pull can be referred to as item difficulty, with some pictures resulting, on average, in more motive-related content than others (as confirmed by, for instance, Pang & Schultheiss, 2005; Schultheiss & Brunstein; 2001; Siefert et al., 2016; Stein et al., 2014). Phrased more generally, card pull can be understood as "a confirmation of Murray's (1943) early assumption that selected TAT images can highlight different psychological themes or emotions" (Auletta et al., 2020, p. 1368). This is also attested by historic accounts on the development of the TAT which state that

> each TAT story told about each card was examined and a rating was given to each card corresponding to the amount of information [about motives] it contributed [. . .]. The average of the ratings given to each card thus reflected its 'stimulating power.' Presumably, the cards with the greatest stimulating power were selected. (W. G. Morgan, 1995, p. 237)

In other words, test development favored pictures (cards) with the greatest stimulating power. Notably, Pang and Schultheiss (2005) revealed that the extent of motive imagery elicited by picture cues showed a pattern that seemed to be consistent across studies. To arrive at this conclusion, these authors compared results from a U.S. sample with those obtained from a German sample (Schultheiss & Brunstein, 2001). Interestingly, the same cards were identified as "high pull" in both samples. Thus, Pang and Schultheiss (2005) stated that "pictures have specific motivational signatures that are robust and can be replicated across cultures" (p. 288). Hence, according to this view, PSE picture cues arouse motives. Note, however, that a comparison of pictures with regard to the resulting motive-related content does not present strong evidence for the causal assumption inherent in this view of how PSEs work (i.e., that pictures arouse motives). As one of many possible alternative explanations, pictures may vary in the types of stories they elicit and the motive content may only be a covariate of the chosen stories.

Pictures Only Eliciting Stories. Research on human motivation unanimously agrees that a vast array of contextual variables may elicit goal-directed behavior (for an overview, see Eccles & Wigfield, 2002). It follows that, while taking a PSE, participants' motives can be aroused by many variables beyond the presented pictures. In fact, early experimental studies by Atkinson and McClelland (1948) revealed that deprivation of basic motives, for example food, can even result in higher motive-related, that is, food-related, content in PSEs. Across seven TAT pictures, these authors found that the percentage of food-related content in stories told, as a response to pictures that are mostly nonfood related, was a function of the hours subjects had been deprived from food (1 vs. 4 vs. 16 hours).

Although one may argue that Atkinson and McClelland (1948) created rather strong contextual influences by depriving participants from food for as long as 16 hours, other theorizing and empirical evidence suggests that subtle influences can also determine motive-related content in PSE stories. The dynamic apperception theory on motivation (Atkinson & Birch, 1970) posits that motive strength changes dynamically and that motives are constantly competing with each other to be expressed in behavior. Thus, motive arousal is also always related to the circumstances of a person (e.g., gender, age, career aspirations, living conditions; see, e.g., Jenkins, 1987; Veroff et al., 1984). Building on this theorizing and assuming that telling an imaginative story is an act of motive satisfaction itself (see McClelland, 1980), Lang (2014) revealed that dynamic item response modelling described their PSE data more adequately than conventional item response approaches. Hence, the previously told story may be a stronger determinant of the current story than the currently shown picture. Some authors even noted that motives can be constantly ("chronically") aroused (Schultheiss et al., 2010).

Ultimately, several authors noted that any free speech can be used to detect implicit motives of the speaker, notwithstanding the absence of specific cues that might arouse the motive. For instance, McClelland (1987) argued motives can be best observed in dreams, fantasies, or free associations. This is echoed by Winter (1994), who developed a scoring system to detect motives in running text. This system was, for instance, used to uncover motives of political leaders on the basis of speeches they gave (Winter et al., 1991). This, together with the aforementioned theorizing and evidence, suggests that it may suffice for PSEs to produce stories that evoke motives without the need to present pictures beforehand (PSEs are then a misnomer). In other words, ambiguous pictures may function as triggers of imaginative stories (since people typically do not simply report their fantasies), but do not necessarily have to arouse specific motives.

Research on card pull indicated that the relative importance of picture imagery in comparison with story content is less important as previously expected (Jenkins et al., 2020; Siefert et al., 2016; Stein et al., 2014). Although research on card pull revealed important insights into the differences between cards in the amount of motive content that is typically produced by test takers, it did not directly tackle the question of whether pictures arouse motive *and* elicit stories or *only* elicit stories. This may also be due to some of the previous literature attributing both functionalities to picture cues (arousing motives and eliciting stories) as if they were the same (Schultheiss & Pang, 2007; Weiner & Greene, 2017).

To shed light on this specific issue, a first piece of experimental evidence comes from research on a *semi*projective test (Krumm et al., 2016). Unlike projective tests, semiprojective tests do not ask test takers to write stories. Instead, test takers go through several (motive-related) statements (e.g., "one might meet someone here" as a statement addressing affiliation), which are presented below each picture, and they check boxes for statements that, in their opinion, apply to the picture at hand. Krumm et al. manipulated the presentation of the stimuli (i.e., the pictures) by presenting items either with or without pictures. Scores are obtained by counting the number of checked statements that refer to a particular motive. Although semiprojective tests share the same rationale as projective tests in that pictures are designed to arouse motives, Krumm et al. revealed that—for the specific test they investigated—3 out of 6 motive scores did not differ regardless of the test was presented whether with or without pictures. These authors concluded that pictures may not be essential components of semiprojective test; in other words, pictures may not causally arouse motive-related responses. Considering this evidence as well as the above theorizing, we preregistered and tested the following hypothesis:

> There will be no significant differences in motive scores resulting from a PSE administered with versus without pictures.[1]

Present Study

To examine the role of pictures in PSEs, the present study adopted an experimental test validation approach (Bornstein, 2011; Borsboom et al., 2004; Erdfelder & Musch, 2006; Krumm et al., 2017). That is, we manipulated a crucial element of a test and examined whether it made a difference for test results. In a way, this is similar to the experimental approaches by Atkinson and McClelland (1948). That is, we manipulated the presumably motive-arousing feature of the test: We either presented ambiguous pictures or repeatedly presented a neutral condition (identical to the white card in the TAT). We randomly assigned participants to one

of the two experimental conditions. If groups differ with regard to the outcome variable in question (in our case: motive scores), a causal effect of the independent variable (in our case: pictures) is evident. Although this is a standard scientific procedure, psychometric test scores are rarely validated by means of experiments (Bornstein, 2011). That is why this experimental validation approach has been strongly recommended (e.g., Borsboom et al., 2004; Krumm et al., 2017). By applying this approach to a classic PSE, we seek to contribute to further clarify the role of pictures for PSEs.

Method

Sample

An a priori power-analysis (G*Power; Faul et al., 2007) revealed that $N = 278$ participants would be required to detect small differences between both groups with sufficient statistical power ($1 - \beta = .80$; assumed effect size of $d = .30$). We chose a small effect size of $d = .30$ to be conservative about the required sample size. In an online study, we tested our hypothesis in a sample of 281 participants (64.8% female, $M_{age} = 29.31$ years, $SD_{age} = 10.23$). Among these, 62.7% of the sample were students and 32.5% were working people. Furthermore, we found a wide variety of educational levels: That is, the majority of the sample held a university entry qualification (A-level, 52.3%), 24,6% held a university bachelor or master's degree and 12.1% held a 10th-grade degree.

Participants were recruited via two different channels: One part of the participants was made aware of the study via online postings (e.g., student and local Facebook-groups [e.g., "sharing is caring"] and university e-mail newsletters; 105 test takers in the condition with pictures and 83 in the condition without pictures) and the other part (33% of the sample; 47 test takers in the with picture condition and 46 in the without picture condition) was surveyed by an online panel that consisted of more than 600,000 test takers who had declared their interest in online surveys (available through https://www.testingtime.com). The panel client base consists of both companies and universities.[2] University students majoring in psychology received course credit for participation; panel participants received 7 €. Furthermore, all participants received feedback on their Big Five personality dimensions.[3]

Study Design and Materials

All data were collected online and followed the recommendations for PSE online administration by Gruber and Kreuzpointner (2015). Participants were randomly assigned to either one of two versions of the PSE. In the condition with pictures, participants saw pictures taken from the

original TAT picture set (Cards 1, 2, 4, 10, and 13MF[4]) for 20 seconds each and were after each picture prompted to write a story, which should answer the following questions:

> 1. What is happening? / Who are the persons? 2. What led up to this situation? / What happened in the past? 3. What is being thought and felt? / What do the persons want? 4. What will happen? / What will be done? (see Murray, 1943; Weiner & Greene, 2017).

In the condition without pictures, participants were asked five times to write a story and received that same four prompts, but never saw a picture (see the appendix for instructions in both conditions). In both conditions, participants had 5 minutes to write down their story (see Weiner & Greene, 2017). We followed the recommendations by Schultheiss and Pang (2007) who recommended using at least four but less than eight picture cues (to avoid fatigue; for a study with a similar number of pictures see, e.g., Schultheiss & Brunstein, 2001). To make the selection of the pictures as representative as possible, we selected five out of the 10 most frequently used TAT pictures (Keiser & Prather, 1990).

We hired two independent, trained coders who were not aware of the purpose of this study and did not know the experimental conditions under which the stories were written. These coders rated all stories for their motive-related content using an established coding manual (Winter, 1994). They provided scores for three broad motive domains (achievement, power, and affiliation) and 15 subscores for narrow aspects of the motive domains (for the list of subscores see Table 1). Interrater reliability of their codings was assessed with intraclass correlation coefficients (ICCs) using two-way mixed effect ICC models. According to recommendations of Shrout and Fleiss (1979), we found excellent agreement, both in the condition with pictures (ICC [3, k] = .89) and in the condition without pictures (ICC [3, k] = .90). We used the average score of both coders for further analyses (see Schultheiss & Pang, 2007). To control for word counts, we used residual z-scores in all subsequent analyses. That is, we used residual z-scores by applying a regression analysis to residualize motive scores for word count (see Schultheiss & Pang, 2007). Internal consistency was generally low (ranging from $\alpha = -.07$ to .38). These estimates are in line with previous findings (see Schultheiss et al., 2008) and, importantly, did not significantly differ between both conditions (all p values > .05, for further information regarding the test for differences between alphas, see Feldt et al., 1987).

After completing the PSE (either with or without pictures), participants worked on the short version of the Big Five Inventory (Rammstedt & John, 2005). Hence, individuals answered 21 items on a 5-point Likert-type scale

(ranging from 1 = *disagree strongly* to 5 = *agree strongly*). Internal consistency of this measure's ratings was acceptable ($\alpha = .63$) to good ($\alpha = .84$).

Furthermore, we tested whether test-taking motivation differed between the conditions with versus without pictures. Therefore, test takers completed five items of the Test Attitude Survey (Arvey et al., 1990) and responded on a 5-point Likert-type scale from *disagree strongly* (1) to *agree strongly* (5). The reliability of this measure's ratings was good ($\alpha = .80$).

Results

Preliminary Analyses

Before testing our main Hypothesis, we checked whether the random assignment had worked. That is, we tested whether the experimental conditions differed on demographics or personality. Both groups (with vs. without picture) did not differ in terms of gender, $\chi^2(2) = 5.957$, $p = .05$, $\phi = .146$; age, $t(279) = 0.379$, $p = .71$; occupation, $\chi^2(5) = 1.124$, $p = .95$, $\phi = .063$; level of education, $\chi^2(9) = 7.418$, $p = .59$, $\phi = .162$; or Big Five personality trait ratings, $F(5, 275) = 0.232$, $p = .95$, Wilk's $\Lambda = 0.996$, partial $\eta^2 = .004$.

Hypothesis Tests

Average story length per picture was 102.29 and 117.47 words, respectively, in the conditions with and without pictures. Interestingly, participants wrote significantly longer stories in the condition without pictures, $t(279) = 3.458$, $p = .001$, Cohen's $d = .414$. Note, however, that the herein used scores were corrected for word count (as recommended by Schultheiss & Pang, 2007). Means and standard deviations of the scores obtained for the two experimental groups are given in Tables 1 and 2. The largest motive score was observed for affiliation when assessed in the condition with pictures. Interestingly, the second largest score was also obtained for the affiliation motive, but in the condition without pictures. In both conditions, mean motive scores differed substantially across motive domains, with achievement motive scores being lowest in both conditions.

To test our Hypothesis, we examined differences in motive scores between both versions of the PSE (with and without pictures). Thus, we conducted a one-way multivariate analysis of variance with the experimental condition (PSE with vs. without pictures) as independent variable and the three motive scores as dependent variables.[5] The omnibus multivariate analysis of variance test revealed a significant main effect, $F(3, 277) = 23.626$, $p < .001$, Wilk's $\Lambda = 0.796$, partial $\eta^2 = .204$, indicating that the availability of pictures had a significant impact on motive scores. Further analyses revealed that the experimental manipulation had

Table 1. Means and Standard Deviations for TAT Subscores in the Two Experimental Conditions.

TAT subscores	Description	Condition with pictures		Condition without pictures		Cohen's d
		M	SD	M	SD	
Affiliation	Expression of positive, friendly, or intimate feelings toward others	8.74	4.82	4.37	4.33	0.84 [0.59, 1.08]
	Negative feeling about separation or disruption of a friendly relationship	2.19	2.58	0.72	1.14	0.66 [0.41, 0.89]
	Affiliative, companionate activities	2.20	1.99	3.86	4.26	−0.43 [−0.67, −0.19]
	Friendly nurturant acts	1.95	3.41	0.98	1.39	0.43 [0.19, 0.66]
Achievement	Adjectives that positively evaluate performance (e.g., "good" or "best")	0.16	0.66	0.27	0.79	−0.09 [−0.32, 0.15]
	Goals/performances that are described in ways that suggest positive evaluation	1.47	1.85	1.34	1.79	0.00 [−0.23, 0.23]
	Mention of winning or competing with others	0.04	0.23	0.41	1.38	−0.39 [−0.63, −0.16]
	Failure, doing badly, or other lack of excellence	0.33	0.80	0.19	0.51	0.21 [−0.03, 0.44]
	Unique accomplishment	0.01	0.11	0.04	0.20	−0.16 [−0.40, 0.07]
Power	Forceful actions which inherently have impact on other people or the world	1.53	1.90	0.92	1.64	0.30 [0.06, 0.54]
	Control or regulation	1.74	1.95	0.45	1.00	0.86 [0.61, 1.10]
	Attempts to influence, persuade, convince, make or prove a point	0.50	0.93	0.19	0.55	0.34 [0.10, 0.58]
	Giving help, advice, or support that is not explicitly solicited	0.49	0.90	0.54	0.87	0.00 [−0.24, 0.23]
	Impressing others or the world at large: mention of fame, prestige, and reputation	0.92	1.28	0.56	1.08	0.35 [0.11, 0.59]
	Strong emotional reaction in one person to the action of another person	0.90	1.26	0.78	1.10	0.02 [−0.21, 0.26]

Note. $n_{with\ picture}$ = 152, $n_{without\ picture}$ = 129; 95% confidence intervals in brackets. Higher effect sizes reflect a higher motive score on facets with pictures compared with facets without pictures. TAT = Thematic Apperception Test.

Table 2. Means and Standard Deviations in the Two Experimental Conditions.

Motive domains	Condition: with pictures		Condition: without pictures	
	M	*SD*	*M*	*SD*
Affiliation	15.07	6.11	9.92	6.94
Achievement	2.01	2.18	2.24	2.99
Power	6.06	3.81	3.44	3.20

Note. Scores per 1,000 words. Mean word count per participant: with pictures = 511 (*SD* = 160.61), without pictures = 587 (*SD* = 207.17). $n_{\text{with picture}}$ = 152, $n_{\text{without picture}}$ = 129.

medium to large effects on affiliation (Cohen's *d* = 0.64, 95% confidence interval [CI: 0.40, 0.88]) and on power motive scores (Cohen's *d* = 0.67, 95% CI [0.43, 0.91]), which were both higher in the condition with pictures than in the condition without pictures, $F(1, 279) = 28.794$, $p < .001$, partial $\eta^2 = .094$, and $F(1, 279) = 31.119$, $p < .001$, partial $\eta^2 = .100$, respectively. However, the achievement motive score was not affected by the availability of pictures, $F(1, 279) = 1.160$, $p = .28$, partial $\eta^2 = .004$. In fact, the motive score in the condition without pictures was even slightly higher than in the condition with pictures (Cohen's *d* = −0.13, 95% CI[−0.36, 0.11]). In sum, our hypothesis was not supported for two out of three global motive scores.

We also inspected mean differences across the two experimental conditions on the level of subscores (see Table 1). We found effect sizes with CIs including zero or even pointing in the opposite direction (i.e., higher motive scores in the condition without pictures) for more than half of the subscores (8 out of 15). That is, for about 53% of the subscores it did not make a difference whether the pictures—the core feature of a PSE—were presented or not. Please note that certain subscores had a low frequency in both conditions. Thus, large sample sizes are required to detect substantial differences for these subscores.

In a next step, we conducted analyses to investigate whether omitting pictures in the PSE changed its correlation with other variables, that is, with broad personality (Big Five) domains. In most prior research, PSEs typically do not show a lot of overlap with traditional self-reports of personality (Köllner & Schultheiss, 2014). So, we verified whether this was still the case when the PSE was administered without pictures. In other words, we checked whether omitting pictures might have shifted the PSE to a measure of explicit motives. In line with previous findings, we found mostly small correlations between PSE scores and personality ratings. Table 3 shows that omitting pictures had almost no effect on correlations with Big Five dimensions. This is also attested by an average difference between correlations of $|\Delta r| = .08$. However, agreeableness showed a higher correlation with affiliation in the condition without pictures than in the condition with pictures (*r*s = .24 and .09, respectively), but this difference in correlations is not significant. Finally,

Table 3. Correlations of TAT Motive Scores with Broad Personality Domains.

Motive domains	O	C	E	A	N
With pictures					
Affiliation	−.14	.00	.12	.09	−.04
Achievement	.05	−.04	.04	.02	−.02
Power	−.05	−.12	.09	−.04	.03
Without pictures					
Affiliation	−.13	−.01	.04	.24**	.03
Achievement	−.09	.00	.01	.03	−.13
Power	−.06	.03	−.16	−.09	.11

Note. $n_{\text{with picture}}$ = 152, $n_{\text{without picture}}$ = 129. Mean word count per participant: with pictures = 511 (*SD* = 160.61), without pictures = 587 (*SD* = 207.17). TAT = Thematic Apperception Test; O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism.
*$p < .05$. **$p < .01$.
All data and code are available on the Open Science Framework (https://osf.io/r57ke/)

we tested whether omitting pictures from a PSE led to a reduction in test-taking motivation. However, results revealed no differences between both groups, $t(279) = 0.543$, $p = .59$, Cohen's *d* = .065, 95% CI[−0.17, 0.30].

Discussion

Theory and research on the assessment of human motives through PSEs make two different assumptions about the role of picture cues. That is, pictures may either be considered to arouse motives and elicit stories or to only elicit stories. By employing an experimental test validation approach (e.g., Bornstein, 2011), the current study made several contributions to disentangle these issues.

First, we found inconsistent effects for the Big-Three motive domains. For the affiliation and the power motive, mean scores of the condition with pictures significantly exceeded those of the condition without pictures. We thus conclude that the chosen pictures did not only elicit stories but also aroused the affiliation and power motives. However, results for the achievement motive lead to the opposite conclusion. Mean achievement motive scores did not differ

significantly across conditions, suggesting that the herein chosen pictures did not function as arousing elements of this motive. In other words, the pictures chosen in the current study, which are frequently used in other TAT studies, elicited stories that contained content unrelated to the achievement motive domain (Schultheiss & Brunstein, 2001; Tuerlinckx et al., 2002). Importantly, for all Big-Three motive domains, we found motive scores above zero in the condition without pictures. This means that ambiguous pictures may function as triggers or amplifiers of imaginative stories; however, they do not seem mandatory to arouse a specific motive. It seems that it may suffice to produce stories that evoke motives (albeit to a lesser degree) without the need to present pictures beforehand. This is in contrast to the notion that PSEs are stimulus tests (Bornstein, 2011) but in line with previous research on picture imagery that found that the relative importance of picture imagery in comparison with story content is less important as previously expected (Siefert et al., 2016; Stein et al., 2014).

The inconsistency of the results across motive domains is, in fact, in line with similar research on a semiprojective test, which also yielded significant effects of omitting picture for the power and affiliation motives (but only their fear component), but not for the achievement motive (Krumm et al., 2016). In light of the currently available evidence, the question about the role of pictures in PSEs—arouse motive and elicit stories versus only elicit stories—might thus be answered with: It depends on the motive domain and the chosen pictures.

Second, we highlight the importance of disentangling the effects of pictures in PSEs. As noted above, several authors seem to use the phrases "elicit stories" and "arouse motives" interchangeably (Schultheiss & Pang, 2007; Weiner & Greene, 2017). We agree that pictures in PSEs might indeed have both effects on test takers, which is also supported by our data. Nevertheless, we posit that it is important to distinguish these two effects, not only to provide more detailed knowledge about existing PSEs, but importantly also to refine the development of new PSEs. If test developers simply want their pictures to elicit imaginative stories, it may not be much of a concern that "less attention has been paid to the specification and selection of picture cues" (Schultheiss & Brunstein, 2001, p. 72). On the other hand, if pictures are implemented in a PSE to not only elicit stories but to specifically arouse motives, the motivational signature of each single picture needs to be taken into account very carefully (Pang & Schultheiss, 2005).

A precise specification of the intended effect of pictures is also important for the psychometric evaluation of PSEs. A dynamic Thurstonian item response approach, as adopted by Lang (2014) to assess the reliability of PSEs, essentially assumes that pictures elicit stories but that the motive arousal and satisfaction is in constant flow. Such reliability estimates are thus incompatible with PSEs in which each picture is designed as a discrete cue to arouse a motive. We therefore suggest that test authors specify the intended effect of pictures and present empirical evidence that is aligned to the specified effect. A potential starting point may be derived from the stimulus sampling from suggested by S. R. Jenkins (2017). She argued that pictures may be systematically sampled to depict (a) plausible role relationships of the people, (b) the dominant activity, and (c) affective tone. Depending on the assessment purpose, test administrators may consider these dimensions more or less relevant and sample pictures accordingly.

Third, our results show that the mean motive content and, even more importantly for assessment purposes, interindividual variability in motive content can be substantial—even when no pictures are presented. However, this result differed drastically across motives. Although a mean affiliation score around 10 was observed in stories told after seeing a blank screen (no picture condition), the mean score was only about 2 for the achievement motive. This pattern may be specific to our sample, the chosen cards, and the study setting; it is difficult to explain why such a pattern emerged. However, marked differences between motive scores in a PSE without pictures bear an important conclusion for research on card pull (e.g., Siefert et al., 2016). This line of research has so far, to our knowledge, compared motive scores across pictures and, on this basis, has drawn inference about the "pull" of pictures for specific themes and personality dimensions (e.g., Cramer, 2017). Ignoring that what one may call the "baseline level" of motive imagery can vary across motives might lead to false conclusions. Thus, we suggest that future research on the pull of individual cards includes information on the baseline level of motive expression, that is, the level of motives expressed without pictures.

Interestingly, the overall finding that high scores in the condition without pictures were associated with even higher scores in the condition with pictures is consistent with the stochastic drop-out apperception theory (Tuerlinckx et al., 2002). This theory assumes that responses to PSEs can be described in two stages. First, a picture may appeal to a motive or not. If it does, the response will contain motive-related content reflecting the motive disposition of a test taker. If it does not, stories will contain mostly irrelevant material. Transferring this reasoning to our results, it may have been rather easy in our sample to appeal to the affiliation motive given the high score in the condition without pictures. Consequently, presenting pictures resulted in even higher scores. As another explanation, it may have been rather difficult to appeal to the achievement motive in our sample. Assuming that most pictures did not appeal sufficiently to the achievement motive, it makes sense that achievement motive scores were not different in both conditions.

Fourth, it may be tentatively concluded that individual differences in the Big-Three motives can be gauged regardless of the presence or absence of pictures. Two results speak to this conclusion: the identical standard deviations and only very small differences in correlations with Big Five domains. Standard deviations were almost identical across conditions. Thus, omitting pictures did not result in ceiling or bottom effect or narrow interindividual differences in any other way. Relatedly, correlations of motive scores with Big Five domains were small, which is in line with previous research (e.g., Pang & Schultheiss, 2005) and, more importantly, were similar across both experimental conditions. This may be viewed as further evidence that interindividual differences in motive scores derived from imaginative stories only, i.e., without being specifically aroused through pictures, may reflect valid interpretations (see also McClelland, 1987; Winter, 1994). However, our findings are in contrast to Jenkins et al. (2020) who used generalizability theory and uncovered that a significant proportion of variance was due to person-card interactions. Hence, future research should examine the effects of presenting versus omitting picture cues on the construct-related validity of PSEs in more detail. Moreover, we encourage future research to examine the effects of presenting versus omitting picture cures on PSEs' criterion-related validity.

From a more practical perspective, investigating the effects of pictures on PSE responses is of importance to test developers. Viewing ambiguous pictures as mere vehicles to get imaginative stories out of test takers means that much less effort needs to be invested into picture design and selection. On the contrary, when test developers are keen on eliciting stories *and* arousing motives, fine-grained knowledge about picture cues and their combination with other pictures is needed. As mentioned above, either one of the two viewpoints also calls for different psychometric approaches. Given that more research is needed, we refrain from presenting concrete practical recommendations other than that test developers be specific about their intended role of pictures. One way of testing the impact of a picture cue provides the experimental test validation approach presented here (see Borsboom et al., 2004; Erdfelder & Musch, 2006; Krumm et al., 2017; for further examples, see Krumm et al., 2016; Schäpers et al., 2020).

In terms of limitations, we first acknowledge that our results are based on a selection of five pictures. Our main criteria for picture selection were (a) to take pictures from the TAT as a classic PSE and (b) to randomly chose five from the ten most frequently used pictures as reported, for example, by Keiser and Prather (1990). In doing so, we sought to come up with a picture set that was representative of the pictures used in research and practice, while at the same time avoiding a biased selection process which could potentially lead to a picture set favoring our hypothesis. However, we acknowledge that the randomly created picture set may not meet all criteria for optimal picture selection as delineated by several authors (Schultheiss & Pang, 2007; Smith et al., 1992). According to these authors, researchers must carefully consider the pictures' content, their ability to pull motives, and their ambiguity as well as the number of pictures.[6] Concerning the number of pictures, we followed Schultheiss and Pang's recommendation to include a minimum of four pictures. Regarding the content of pictures, we acknowledge that the depicted actors may be viewed as being from another time and thus out of date, which may affect participants' stories (as cautioned by Smith et al., 1992). However, this should not have affected only the achievement-related story content in the current study. Card pull does indeed differ across TAT cards (Siefert et al., 2016). Interestingly, a review by Stein et al. (2016), which included the herein used Cards 1, 2, 4, and 13MF, suggests that these cards can be expected to pull the Big-Three motive domains. For instance, achievement is among the most frequently occurring topics in responses to Cards 1 and 2, whereas power-related content seems to be frequently occurring in responses to Cards 4 and 13MF. Moreover, Stein et al.'s review reveals that those cards elicit a variety of different topics, thus meeting the requirement for pictures to be sufficiently ambiguous (Schultheiss & Pang, 2007). Nevertheless, we acknowledge that the current findings are based on a particular selection of pictures and need replication with different TAT cards as well as with other PSEs, which come with ample further pictorial material that was not subject to this study (e.g., Runge & Lang, 2019).

Second, the substantial differences in motive content, which we observed in the condition without pictures, may represent a specificity of our sample. Notwithstanding the absence of differences in motive arousal among our subsamples (as defined by different recruiting strategies) and other studies confirming similar card pulls across cultures (Pang & Schultheiss, 2005), further studies are needed to examine the generalizability of this finding. In particular, considering that external events (e.g., the COVID-19 pandemic) might also have an impact on test takers' motive arousal, future research is required to examine the generalizability of our results (see also Veroff et al., 1984).

Third, we chose an online test environment. As outlined by Aronow et al. (2001), test environment may affect PSE responses. A number of researchers suggested that a test-administration by a human experimenter could lead to different implicit motive-scores than a computer-based test-situation (see Gruber & Kreuzpointner, 2015). One might assume that characteristics (e.g., status of the person) or nonverbal behavior of the test administrator have an impact on the stories developed in the test situation (e.g., Klinger, 1967). Thus, we recommend additional research in a proctored setting. Note, however, that Gruber and Kreuzpointner (2015) as well as Bernecker and Job (2011) confirmed the feasibility and robustness of PSE online

administration. In fact, Bernecker and Job (2011, p. 262) concluded as follows:

> Stories written online turned out to be denser in motive imagery then stories written in the lab, particularly in affiliation and achievement imagery. One reason could be that the setting which most participants indicated to be a "private place" allowed them to fantasize more freely.

Conclusion

This study draws a specific distinction between pictures in PSEs as either eliciting stories or arousing motives. On the basis of an experimental test validation approach, we conclude that the chosen pictures aroused affiliation and power motives, but not the achievement motive. We therefore suggest that test authors specify the intended effect of pictures and present empirical evidence that is aligned to the specified effect. Moreover, we encourage more research on the impact of pictures on the construct-related validity of PSEs.

Appendix

General Participant Instructions (Condition With Pictures)

You will now see several pictures. Your task is to tell a story to each one of these pictures. Try to imagine what might be happening on each picture. Please tell us, what the situation is like, how things led to this situation, what the acting persons think and feel and what they might do next. In other words: Write a proper story with a plot and with characters. You will be given 5 minutes for each story and you will be told when it is time to end your story and prepare for the next picture and the next story. Write five different stories. There is no correct or incorrect story or type of story, so feel free to write any story that comes to your mind when looking at the picture. Before you can write a story, each picture will be shown to you for 20 seconds. After that, the picture disappears and you can start writing the story. To ensure that you have devoted enough time to each story, you will only be allowed to finish a story and move on to the next picture, by clicking on "next," after 4 minutes. Please insert the digit five in the box below, so we know you have read and understood the instructions. After that, please click on "next."

General Participant Instructions (Condition Without Pictures)

Your task is to tell a story. Try to imagine what might be happening. Please tell us, what the situation is like, how things led to this situation, what the acting persons think and feel and what they might do next. In other words: Write a proper story with a plot and with characters. You will be given 5 minutes for each story and you will be told when it

is time to end your story and prepare for the next story. Write five different stories. There is no correct or incorrect story or type of story, so feel free to write any story that comes to your mind. Before you can write a story, we will ask you to think about the story you want to write for 20 seconds. After that, you can start writing the story. To ensure that you have devoted enough time to each story, you will only be allowed to finish a story and move on, by clicking on "next," after 4 minutes. Please insert the digit five in the box below, so we know you have read and understood the instructions. After that, please click on "next."

Stimulus Material and Specific Instruction (Condition With Pictures)

A picture and a timer (counting from 20 seconds to 0) are presented along with the instruction: "Watch the picture for 20 seconds"

Stimulus Material and Specific Instruction (Condition Without Pictures)

A timer (counting from 20 seconds to 0) is presented along with the instruction: "Think about the story you want to write for 20 seconds"

Subsequent Prompts (Identical in Both Conditions)

"Write a story.

ORCID iD

Philipp Schäpers [iD] https://orcid.org/0000-0002-8270-5105

Notes

1. Hypothesis and research design are preregistered at https://osf.io/r57ke/
2. Due to the different recruitment strategies, we collected data that mainly came from students (online postings) and data that is dominated by working people (online panel). Thus, both groups (online postings vs. online panel) differed in gender, $\chi^2(2) = 11.33$, $p < .05$, $\phi = .201$, age, $t(279) = 5.378$, $p < .01$; occupation, $\chi^2(5) = 72.718$, $p < .01$, $\phi = .509$; and level of education, $\chi^2(9) = 51.976$, $p < .01$, $\phi = .43$. Importantly, we did not find any differences in Big Five personality, $F(5, 275) = 1.793$, $p = .11$, Wilk's $\Lambda = 0.968$, partial $\eta^2 = .032$, or motive imagery; with picture: $F(3, 148) = 2.112$, $p = .10$, Wilk's $\Lambda = 0.959$, partial $\eta^2 = .041$; without picture: $F(3, 125) = 0.478$, $p = .70$, Wilk's $\Lambda = 0.989$, partial $\eta^2 = .011$, between both groups.
3. We followed recommendations by Meade and Craig (2012) and added two bogus items and a self-declaration of data exclusion to detect careless responding. Of the initial sample of 295 participants, 14 were excluded because they failed the bogus items or self-declared to be better excluded from further analyses.
4. "The original TAT cards are numbered from 1 to 20, and nine of the cards are additionally designated by letters intended to indicate their appropriateness for boys (B) and girls (G) ages 4 to 14 years, males (M) and females (F) ages 15 years or older, or some combination of these characteristics (as in 3BM, 6GF, 12BG, and 13MF)" (Weiner & Greene, 2017, p. 391).
5. We are aware that assuming a nonsignificant result in a hypothesis is usually followed up by Bayesian analyses (e.g., Wagenmakers et al., 2018). In order to keep analyses as straightforward as possible as well as comparable to previous research on card pull (e.g., Schultheiss & Brunstein, 2001), we decided to report results based on classical null hypothesis significance testing. However, we found similar results when conducting a Bayesian approach. Details about the Bayesian analyses can be requested from the first author.
6. Note that Smith et al. (1992) also added the order of pictures as a relevant desideratum. However, we refer to Schultheiss and Pang's more recent conclusion that the sequence of pictures has only a marginal effect on motive expression. Furthermore, Schultheiss and Pang suggested to use pictures which are somewhat similar to the criterion to be predicted (e.g., a picture showing a ship captain to predict participants' persuasiveness). Since no criteria were included our study, we do not discuss this aspect of picture selection here.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Aronow, E., Altman Weiss, K., & Reznikoff, M. (2001). *A practical guide to the Thematic Apperception Test: The TAT in clinical practice*. Brunner-Routledge.

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43*(4), 695-716. https://doi.org/10.1111/j.1744-6570.1990.tb00679.x

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. Wiley.

Atkinson, J. W., & McClelland, D. C. (1948). The projective expression of needs: The effect of different intensities of the hunger drive on thematic apperception. *Journal of Experimental Psychology*, *38*(6), 643-658. https://doi.org/10.1037/h0061442

Auletta, A. F., Cupellaro, S., Abbate, L., Aiello, E., Cornacchia, P., Norcia, C., & Sogos, C. (2020). SCORS-G and card pull effect of TAT stories: A study with a nonclinical sample of children. *Assessment*, *27*(6), 1368-1377. https://doi.org/10.1177/1073191118781014

Baumann, N., Kaschel, R., & Kuhl, J. (2005). Striving for unwanted goals: Stress-dependent discrepancies between explicit and implicit achievement motives reduce subjective well-being and increase psychosomatic symptoms. *Journal of Personality and Social Psychology*, *89*(5), 781-799. https://doi.org/10.1037/0022-3514.89.5.781

Bernecker, K., & Job, V. (2011). Assessing implicit motives with an online version of the picture story exercise. *Motivation and Emotion*, *35*(3), 251-266. https://doi.org/10.1007/s11031-010-9175-8

Blankenship, V., Vega, C. M., Ramos, E., Romero, K., Warren, K., Keenan, K., Rosenow, V., Vasquez, J., & Sullivan, A. (2006). Using the multifaceted Rasch model to improve the TAT/PSE measure of need for achievement. *Journal of Personality Assessment*, *86*(1), 100-114. https://doi.org/10.1207/s15327752jpa8601_11

Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment*, *11*(1), 48-57. https://doi.org/10.1037/1040-3590.11.1.48

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, *23*(2), 532-544. https://doi.org/10.1037/a0022402

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061-1071. https://doi.org/10.1037/0033-295X.111.4.1061

Childs, R. A., & Eyde, L. D. (2002). Assessment training in clinical psychology doctoral programs: What should we teach? What do we teach? *Journal of Personality Assessment*, *78*(1), 130-144. https://doi.org/10.1207/S15327752JPA7801_08

Collins, C. J., Hanges, P. J., & Locke, E. A. (2004). The relationship of achievement motivation to entrepreneurial behavior: A meta-analysis. *Human Performance*, *17*(1), 95-117. https://doi.org/10.1207/S15327043HUP1701_5

Costantino, G., Litman, L., Waxman, R., Dupertuis, D., Pais, E., Rosenzweig, C., Forti, G., Paronik, J., & Canales, M. M. F. (2014). Tell-Me-A-Story (TEMAS) assessment for culturally diverse children and adolescents. *Rorschachiana*, 35*(2)*, 154-175. https://doi.org/10.1027/1192-5604/a000054

Costantino, G., & Malgady, R. G. (2008). Tell-me-a-story (TEMAS). In S. R. Jenkins (Ed.), *Personality and clinical psychology: A handbook of clinical scoring systems for*

thematic apperceptive techniques (pp. 547-572). Lawrence Erlbaum.

Cramer, P. (2017). Defense mechanism card pull in TAT stories. *Journal of Personality Assessment*, *99*(1), 15-24. https://doi.org/10.1080/00223891.2016.1207080

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109-132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Erdfelder, E., & Musch, J. (2006). Experimental methods of psychological assessment. In M. Eid, & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 205-220). American Psychological Association.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, *11*(1), 93-103. https://doi.org/10.1177/014662168701100107

Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology*, *8*(2), 389-413. https://doi.org/10.1080/00223980.1939.9917671

George, C., & West, M. (2001). The development and preliminary validation of a new measure of adult attachment: The Adult Attachment Projective. *Attachment & Human Development*, *3*(1), 30-61. https://doi.org/10.1080/14616730010024771

Gruber, N., & Kreuzpointner, L. (2013). Measuring the reliability of picture story exercises like the TAT. *PLOS ONE*, *8*(11), e79450. https://doi.org/10.1371/journal.pone.0079450

Gruber, N., & Kreuzpointner, L. (2015). Assessing the implicit achievement motive: Effects of input-condition, administration and picture-position. *International Journal of Psychology and Educational Studies*, *2*(3), 10-19. https://doi.org/10.17220/ijpes.2015.03.002

Hibbard, S., Mitchell, D., & Porcerelli, J. (2001). Internal consistency of the object relations and social cognition scales for the Thematic Apperception Test. *Journal of Personality Assessment*, *77*(3), 408-419. https://doi.org/10.1207/S15327752JPA7703_03

Jenkins, R. J. (2017). The Narrative arc of TATs: Introduction to the JPA special section on thematic apperceptive techniques. *Journal of Personality Assessment*, *99*(3), 225-237. https://doi.org/10.1080/00223891.2016.1244066

Jenkins, S. R. (1987). Need for achievement and women's careers over 14 years: Evidence for occupational structure effects. *Journal of Personality and Social Psychology*, *53*(5), 922-932. https://doi.org/10.1037/0022-3514.53.5.922

Jenkins, S. R. (2017). Not your same old story: New rules for thematic apperceptive techniques (TATs). *Journal of Personality Assessment*, *99*(3), 238-253. https://doi.org/10.1080/00223891.2016.1248972

Jenkins, S. R., Siefert, C. J., & Weber, K. (2020). Interpersonal decentering and person–situation interaction in the Thematic Apperception Test: Is It all in the cards? What's the story? *Journal of Personality Assessment*, *102*(4), 551-562. https://doi.org/10.1080/00223891.2018.1557668

Kehr, H. M. (2004). Integrating implicit motives, explicit motives, and perceived abilities: The compensatory model of work motivation and volition. *Academy of Management Review*, *29*(3), 479-499. https://doi.org/10.2307/20159055

Keiser, R. E., & Prather, E. N. (1990). What is the TAT? A review of ten years of research. *Journal of Personality Assessment*, *55*(3-4), 800-803. https://doi.org/10.1080/00223891.1990.9674114

Klinger, E. (1967). Modeling effects on achievement imagery. Journal of Personality and Social Psychology, *7*(1 Pt. 1), 49-62. https://doi.org/10.1037/h0024936

Köllner, M. G., & Schultheiss, O. C. (2014). Meta-analytic evidence of low convergence between implicit and explicit measures of the needs for achievement, affiliation, and power. *Frontiers in Psychology*, *5*, 826. https://doi.org/10.3389/fpsyg.2014.00826

Krumm, S., Hüffmeier, J., & Lievens, F. (2017). Experimental test validation. *European Journal of Psychological Assessment*, *35*(2), 225-232. https://doi.org/10.1027/1015-5759/a000393

Krumm, S., Schäpers, P., & Göbel, A. (2016). Motive arousal without pictures? An experimental validation of a hybrid implicit motive test. *Journal of Personality Assessment*, *98*(5), 514-522. https://doi.org/10.1080/00223891.2016.1146290

Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., & Eisman, E. J. (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology: Research and Practice*, *31*(2), 119-130. https://doi.org/10.1037//0735-7028.31.2.119

Lang, J. W. B. (2014). A dynamic Thurstonian item response theory of motive expression in the picture story exercise: Solving the internal consistency paradox of the PSE. *Psychological Review*, *121*(3), 481-500. https://doi.org/10.1037/a0037011

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Sience in the Public Interest*, *1*(2), 27-66. https://doi.org/10.1111/1529-1006.002

Lundy, A. (1985). The reliability of the Thematic Apperception Test. Journal of Personality Assessment, *49(2)*, 141-145. https://doi.org/10.1207/s15327752jpa4902_6

McClelland, D. C. (1980). Motive dispositions: The merits of operant and respondent measures. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 1, pp. 10-41). Sage.

McClelland, D. C. (1985). How motives, skills, and values determine what people do. *American Psychologist*, *40*(7), 812-825. https://doi.org/10.1037/0003-066X.40.7.812

McClelland, D. C. (1987). *Human motivation*. Press Syndicate of the University of Cambridge.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, *96*(4), 690-702. https://doi.org/10.1037/0033-295X.96.4.690

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437-455. https://doi.org/10.1037/a0028085

Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology & Psychiatry*, *34*(2), 289-306. https://doi.org/10.1001/archneurpsyc.1935.02250200049005

Morgan, W. G. (1995). Origin and history of the Thematic Apperception Test images. *Journal of Personality Assessment*, *65*(2), 237-254. https://doi.org/10.1207/s15327752jpa6502_2

Murray, H. (1943). *Manual for the Thematic Apperception Test*. Harvard University Press.

Pang, J. S. (2010). Content coding methods in implicit motive assessment: Standards of measurement and best practices for the Picture Story Exercise. In O. C. Schultheiss, & J. C. Brunstein (Eds.), *Implicit motives* (pp. 119-150). Oxford University Press.

Pang, J. S., & Schultheiss, O. C. (2005). Assessing implicit motives in U.S. college students: Effects of picture type and position, gender and ethnicity, and cross-cultural comparisons. *Journal of Personality Assessment*, *85*(3), 280-294. https://doi.org/10.1207/s15327752jpa8503_04

Peterson, C. A., & Schilling, K. M. (1983). Card pull in projective testing. *Journal of Personality Assessment*, *47*(3), 265-275. https://doi.org/10.1207/s15327752jpa4703_7

Piotrowski, C. (2017). Thematic Apperception Techniques (TAT, CAT) in assessment: A summary review of 67 survey-based studies of training and professional settings. *SIS Journal of Projective Psychology and Mental Health*, *24*, 11-24. https://www.researchgate.net/publication/312042217_Thematic_Apperception_Techniques_TAT_CAT_in_Assessment_A_Summary_Review_of_67_Survey-based_Studies_of_Training_and_Professional_Settings

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit [Short version of a Big-5 Inventory: Development and validation of an economic inventory to assess the Big-5 personality dimensions]. *Diagnostica*, *51*(4), 195-206. https://doi.org/10.1026/0012-1924.51.4.195

Runge, J. M., & Lang, J. W. (2019). Can people recognize their implicit thoughts? The motive self-categorization test. *Psychological Assessment*, *31*(7), 939-951. https://doi.org/10.1037/pas0000720

Schäpers, P., Mussel, P., Lievens, F., König, C. J., Freudenstein, J.-P., & Krumm, S. (2020). The role of Situations in Situational Judgment Tests: Effects on construct saturation, predictive validity, and applicant perceptions. *Journal of Applied Psychology*, *105*(8), 800-818. https://doi.org/10.1037/apl0000457

Schönbrodt, F. D., Hagemeyer, B., Brandstätter, V., Czikmantori, T., Gröpel, P., Hennecke, M., Israel, L. S. F., Janson, K. T., Kemper, N., Köllner, M. G., Kopp, P. M., Mojzisch, A., Müller-Hotop, R., Prüfer, J., Quirin, M., Scheidemann, B., Schiestel, L., Schulz-Hardt, S., Sust, L. N. N., . . . Schultheiss, O. C. (2020). Measuring implicit motives with the Picture Story Exercise (PSE): Databases of expert-coded german stories, pictures, and updated picture norms. *Journal of Personality Assessment*, *103*(3), 92-405 https://doi.org/10.1080/00223891.2020.1726936

Schultheiss, O. C. (2008). Implicit motives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 603-633). Guilford Press.

Schultheiss, O. C., & Brunstein, J. C. (2001). Assessment of implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment*, *77*(1), 71-86. https://doi.org/10.1207/S15327752JPA7701_05

Schultheiss, O. C., Liening, S. H., & Schad, D. (2008). The reliability of a Picture Story Exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality*, *42*(6), 1560-1571. https://doi.org/10.1016/j.jrp.2008.07.008

Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 322-344). Guilford Press.

Schultheiss, O. C., Rösch, A. G., Rawolle, M., Kordik, A., & Graham, S. (2010). Implicit motives: Current topics and future directions. In T. C. Urdan, & S. A. Karabenick (Eds.), *Advances in motivation and achievement* (Vol. 16a: The decade ahead: Theoretical perspectives on motivation and achievement, pp. 199-233). Emerald.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428. https://doi.org/10.1037//0033-2909.86.2.420

Siefert, C. J., Stein, M. B., Slavin-Mulford, J., Sinclair, S. J., Haggerty, G., & Blais, M. A. (2016). Estimating the effects of Thematic Apperception Test card content on SCORS–G ratings: Replication with a nonclinical sample. *Journal of Personality Assessment*, *98*(6), 598-607. https://doi.org/10.1080/00223891.2016.1167696

Slabbinck, H., De Houwer, J., & Van Kenhove, P. (2013). Convergent, discriminant, and incremental validity of the pictorial attitude implicit association test and the picture story exercise as measures of the implicit power motive. *European Journal of Personality*, *27*(1), 30-38. https://doi.org/10.1002/per.1846

Smith, C. P., Feld, S. C., & Franz, C. E. (1992). *Methodological considerations: Steps in research employing content analysis systems*. In C. P. Smith, J. W. Atkinson, D. C. McClelland, & J. Veroff (Eds.), *Motivation and personality: Handbook of thematic content analysis* (pp. 515-536). Cambridge University Press.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, *112*(1), 140-154. https://doi.org/10.1037//0033-2909.112.1.140

Stein, M. B., Slavin-Mulford, J., Siefert, C. J., Sinclair, S. J., Malone, J. C., Renna, M., Bello, I., & Blais, M. A. (2014). SCORS-G stimulus characteristics of select Thematic Apperception cards. *Journal of Personality Assessment*, *96*(3), 339-349. https://doi.org/10.1080/00223891.2013.823440

Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the Thematic Apperception Test: A psychometric study. *Journal of Personality and Social Psychology*, *82*(3), 448-461. https://doi.org/10.1037/0022-3514.82.3.448

Veroff, J., Reuman, D., & Feld, S. (1984). Motives in American men and women across the adult life span. *Developmental Psychology*, *20*(6), 1142-1158. https://doi.org/10.1037/0012-1649.20.6.1142

Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35-57. https://doi.org/10.3758/s13423-017-1343-3

Weiner, I. B., & Greene, R. L. (2017). Thematic Apperception Test. In I. B. Weiner, & Greene, R. L. (Eds.), *Handbook of personality assessment* (Vol. 2, pp. 391-432). John Wiley.

Winter, D. G. (1994). *Manual for scoring motive imagery in running text* (Version 4.2). University of Michigan. https://deepblue.lib.umich.edu/handle/2027.42/117563

Winter, D. G., Hermann, M. G., Weintraub, W., & Walker, S. G. (1991). The personalities of Bush and Gorbachev measured at a distance: Procedures, portraits, and policy. *Political Psychology*, *12*(2), 215-245. https://doi.org/10.2307/3791463