Content validity of an objective personality test for the assessment of achievement motive

MICHAELA M. WAGNER-MENGHIN[1]

Abstract

Traditionally, variables of achievement motive have been assessed using projective or questionnaire techniques. Under focus here is the test „Work Style", a new technique applying Cattell's objective personality testing approach in the field of achievement motive. „Coding symbols" (CS), a goal setting task and „figure discrimination" (FD), an endurance task are discussed in the framework of achievement motive theory. Using available samples in two different motivational settings (a selection situation and a voluntary research situation), the hypotheses are formulated to test whether achievement motive is activated by the tasks.

The CS-task is currently suitable for evoking achievement motive, which is the precondition for the construction of content valid OPT-variables. First allusions to the construct validity of variables derived from this task are positive. Content validity of the FD-task can not be ensured as the structure of the task is not suitable for deriving valid OPT-variables.

Key words: achievement motive, assessment, objective test, aspiration level, endurance

[1] Mag. Dr. Michaela M. Wagner-Menghin, Division for Assessment and Applied Psychometrics, Division of Psychology, Liebiggasse 5, A-1010 Vienna, Austria; E-mail: michaela.wagner@univie.ac.at

Although the achievement motive is used in a wide sphere of psychological research and practical work areas (e.g. educational psychology) there are only a few psychometric tests quoted to measure this important construct. Within the traditional achievement motive research there are two major approaches to assessing variables of the achievement motive: (1) The story-based measures in Thematic Apperception Test (TAT, McClelland, Atkinson, Clark, & Lowell, 1953; Murray, 1942) have been the standard for assessing the approach component hope for success and the avoidance component fear of failure. For more practical use, especially in the personnel selection context, a projective technique certainly does have certain disadvantages. The recording of answers should be done individually, scoring requires experience with the category system and objectivity is difficult to obtain. Therefore the second approach attempts to replace the TAT with easier to administer and to score questionnaires. However, none of the numerous questionnaires have proved to relate consistently to the story-based motive measures (for an overview see: Schmalt & Sokolowski, 2000). According to McClelland, Köstner and Weinberger (1989) this is due to the nature of the construct. One part of the achievement motive, the self-attributed need achievement, represents the deliberate opinion of a person concerning achievement and can be assessed with questionnaire-type measures. The other part, the implicit need achievement, is unconscious and therefore is only accessible with story-based measures. The two motives have construct validity in different areas. Implicit motives are said to be good predictors of what people do, how they spent their time, and of long-term operant behaviour like career development. Explicit motives predict attitudes, values goals, and effort. Originally conceptualised as independent motive-constructs there is increasing evidence that implicit motives and self-attributed motives are significantly correlated, albeit modestly (Spangler, 1992). Current research focus on identifying factors that influence the degree of concordance between the two constructs, and their power in predicting the adoption of goals and behaviour.

However, this new conception of achievement motive doesn't solve the fundamental problems experienced in assessment. The difficulties in administering and interpreting projective techniques still exist and the use of questionnaires in personnel selection context is still controversial because they are prone to faking (Kubinger, 2002).

Worth mentioning in this context is the recent successful attempt to integrate the advantages concerning validity of story-based measures and the advantages from the administration and scoring using questionnaire-based measures in the semi-projective MMG Multi-Motive-Grid (MMG, Schmalt, Sokolowski, & Langens, 2000). Citing examples from cognitive psychology about functional differences in the processing of verbal and visual information, Schmalt and Sokolowski (2000) analyse the motives activated through the material and the present motives in the process of answering. They deduce that the MMG assesses implicit motives like TAT. Their main argument is that the pictorial material is able to unconsciously stimulate the same process as the real situation does, thus working like a miniature-situation with all emotional and motivational components, whereas lexical material stimulates a more conscious reviewing and evaluation process. This assumption is supported by cognitive psychology's findings that visually induced mental imagery evokes the same processes as in the real situation, visual stimuli also activate semantic memory more quickly than lexical information and visual stimuli activate emotion and motivation relevant memory more quickly and more easily than lexical information does. Looking at the process of answering they deduce that the answer is produced with implicit reference to the motive only in

TAT. Selecting the relevant semantic answer in MMG requires the conscious system, and is thus similar to the explicit process in questionnaires.

Another recent test development concerning the assessment of achievement motive took up Cattell's objective personality testing approach (OPT, Cattell, 1955, 1958; Cattell & Warburton, 1967). The test battery Work Style (Kubinger & Ebenhöh, 1996) aimed to reach the needs of personnel psychology in assessing typical behaviour (style) in achievement situations. OPT measures rely on observing behaviour rather than self-rating behaviour and are therefore claimed to be less prone to faking than questionnaires but more objective than projective measures. Although „Work style" is originally discussed in the context of cognitive styles, the connection to achievement motive is actually given trough variables assessed in subtest 2 and 3: The tasks „Coding of Symbols" (CS) and „Figure Discrimination" (FD) are designed as performance tasks (for a detailed description of these tasks see below). The variables 'Aspiration Level', 'Tolerance for Frustration' and 'Endurance', all constructs discussed in the framework of achievement motive, are defined in the current Work Style manual. Despite of this connection to achievement motive a more detailed discussion of Work Style's tasks and the variables in the framework of achievement motive theory is missing in the manual. This shortcoming concerning content validity and construct validity might be the reason, why the test battery, thought received well by practitioners in personnel psychology, provoked criticism with ongoing use. The problems reported most often regard the norm oriented interpretation of the scored variables which some users mentioned to be inconsistent.

The present study aims to contribute to the validity of the OPT technique in the area of achievement motive. The mentioned Work style tasks and their scores are discussed in the framework of achievement motive theory more detailed than in the manual, to judge about their content validity. Out of this discussion hypotheses are generated and tested by assembling available data into a quasi experimental design to support the content validity of the OPT-technique.

The two tasks

Coding Symbols (CS) is a five trial speed task (Figure 1), based on the paradigm of setting an aspiration level. Additionally repeated experiencing of failure is integrated in the task.

Studies working with the method of setting an aspiration level or setting a goal were originally conducted to prove the validity of the risk-taking model (Atkinson, 1957, 1964). This model combines dispositional and situational factors to derive the resulting motivation. The dispositional factors are the two facets of achievement motive, hope for success and fear of failure, the situational factors are incentives for success and failure and the subjective probabilities of success and failure. Depending on which facet of the achievement motive dominates, the resulting motivation is an approach or an avoiding tendency. Approach and avoidance tendencies are strongest, when the subjective probability of success and failure is about 50% (medium difficult task). Results from studies testing the risk-taking model using the paradigm of setting an aspiration level (Heckhausen, Schmalt, & Schneider, 1985, S. 75ff.) indicate that subjects with higher TAT-hope for success scores set their aspiration level or goals more realistic or slightly optimistic (they reach or nearly reach what they pre-

dict to achieve) than subjects with TAT-scores indicating fear of failure. They tend to set unrealistic (either unrealistic high or unrealistic low) goals.

The OPT-variable „aspiration level" of the CS task is based on these results: It expresses whether the person sets realistic or unrealistic goals. It does not include information about the absolute level of performance, but is individually standardised using the performance of trial 2 as the baseline. The observed aspiration level at the beginning of the test (when there has not been frustration yet) is used as an objective indicator for the dominant type of achievement motive when working on the task. As these variable use the information from the beginning of the CS-task, where the testee only has feedback about his individual performance, it can be labelled more precisely as the „aspiration level under individual standards".

Studies where individuals experience failure are relevant for the interpretation of the OPT-variable „frustration tolerance". Assuming that experiencing failure shifts the subjective probabilities of success and failure the risk-taking model predicts consequences on the strength of the approach and avoidance tendencies, which influence the persistence behav-

Figure 1:
The structure of „Coding Symbols" in diagram form and scored variables
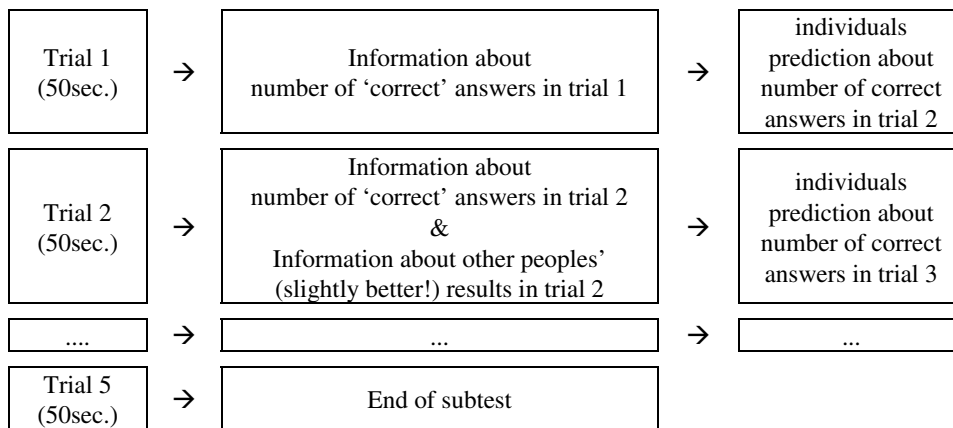
Notes: scored variables according to „Work Style" manual
OPT- variables:
„aspiration level (start of test)" = ('prediction after trial 1' – 'correct symbols in trial 2') / 'correct symbols in trial 2'; (norm oriented interpretation)
„frustration tolerance" = ('prediction after trial 5' - ' prediction after trial 2')/ prediction after trial 2; (norm oriented interpretation)
control variables:
„coding performance" = number of 'correct' answers in trial 2 (norm oriented interpretation);
„trial with maximum performance"
„goal discrepancy" = Sum of absolute discrepancy between trial and forecast.

iour when working on a task. Feather (1961; 1963) studied this relation by inducing failure on repeated tasks and reported supportive results for the risk-taking model. When experiencing failure approach motivated individuals persist longer on easy tasks than avoidance motivated individuals. For difficult tasks the reverse is valid. In a study by Schmalt (1999) endurance in the case of failure, was assessed using unsolvable problems which claimed to be easy or difficult. Given the prospect of proceeding to a medium difficult task, persons with high hope for success scores persisted longer on the current task that they claimed was easy than on the task that they claimed was difficult. In the case of persons with high fear of failure scores there was no difference in persistence.
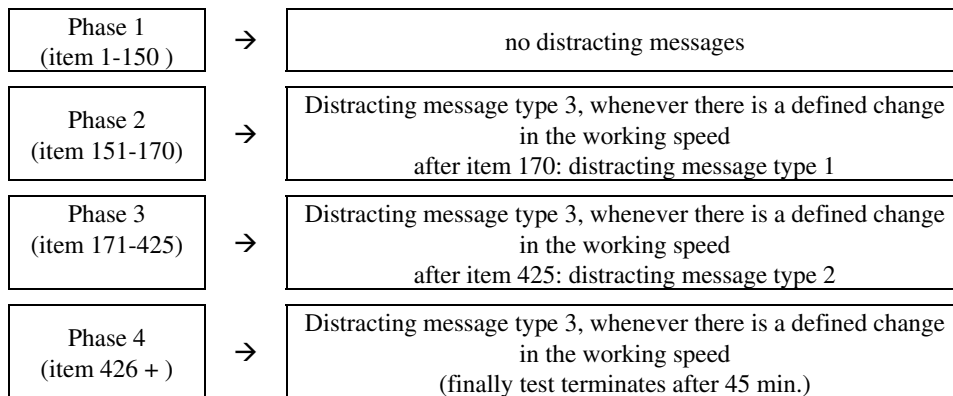
The „Work-Style" OPT-variable which reflects the impact of experienced failure is called „frustration tolerance". Similar to the studies cited above individuals experience repeated failure after trial 2, 3 and 4 in the CS task. Their coding performance is compared to the performance of „others", with the others' performance being slightly better. However the situation is different with respect to the manipulation of the subjective probabilities for success and failure. Individuals in the studies of Feather had a clear idea about success or failure - a puzzle was solved or not. The subjective probability had to be manipulated by the experimenter. Persistence has been measured by using the time spent on task. Individuals working on the CS-task don't have this clear idea about success and failure when starting the task. When doing trial 1 they know only that they have to do as much correct symbols as possible. They have to develop their own criterion for success when they work on trial 2 after they made their prediction, or when experiencing the social comparison. But there is a clear indicator for persistence: A rising (or at least steady) performance over the 5 trials is a sign of persistence on this task. The difficulty of the symbols is constant over the 5 trials and the coding intervals of 50 sec. are short, so there are no effects of tiredness assumed. Performance over the trials 2 to 5 (as well as the prediction) is therefore assumed to rise slightly or to stay constant.

„Frustration tolerance" represents the change in performance between trial 2 and trial 5 and again is individually standardised using the performance of trial 2 as the baseline. A drop in performance is a sign that an individual is not persisting on the task when facing failure. Referring to the studies cited above this can be interpreted as a sign of avoidance motivation.

Figure discrimination (FD) is a power task with self paced termination. The FD-task is obviously easy: in a row of four easy geometrical figures three are identical. The testee has to select which one is different, feedback regarding whether the answer is correct or incorrect is given immediately (the selected field turns green when correct or red when incorrect, any adjustment is then not possible). The subject is free to terminate the task at any time, by selecting a large button „terminate test" in the bottom right corner of the screen. During the course of the test, various messages keep popping up to inform the testee of the possibility of terminating the task at any time (message type 1 and 2) and if necessary about performance irregularities (message type 3). Based on this messages the course of the task can be divided in 4 phases (Figure 2).

Only one variable is scored in the current version of Work Style: It is based on the absolute coding performance and is originally called „achievement motivation". However it is to question, whether this interpretation is a valid one. The current manual does not provide data supporting the interpretation as „achievement motivation" and the discussion regarding content validity is sparse. Citing Heckhausen (1989) the manual refers to McClellands defini-

Figure 2:
The structure of „Figure Discrimination" in diagram form and scored variables

Phase 1 (item 1-150)	→	no distracting messages
Phase 2 (item 151-170)	→	Distracting message type 3, whenever there is a defined change in the working speed after item 170: distracting message type 1
Phase 3 (item 171-425)	→	Distracting message type 3, whenever there is a defined change in the working speed after item 425: distracting message type 2
Phase 4 (item 426 +)	→	Distracting message type 3, whenever there is a defined change in the working speed (finally test terminates after 45 min.)

Notes: 'endurance' = number of correct distracting message type 1: „By working till here you proved as quite successful, but there are other subjects who will still continue". distracting message type 2: „Remember! When you are no longer able to concentrate on the task you have the possibility of terminating it. But consider: There is always somebody continuing working." distraction message type 3 (given when the time between two answers is more than mean of all answer times until now plus 3 times standard deviation of all answer times until now): „You obviously lose your attention"

tion of „achievement motivation". The Work Style authors conclude that they need to define a task, that is rather easy to reduce the influence of „ability", but allows the person to show „endurance" or „effort". The structure of the FD-task makes it plausible that this variable reflects „endurance on easy tasks" and I consider this label as more precise than „achievement motivation". In the current paper this variable is therefore called „endurance".

Possibly corresponding with „endurance" on the task is the indirect influence of the speed component – the slower one works, the longer it takes to reach a disturbing message. Quantifying „endurance" as the number of correct neglects the fact that defining endurance as the „time spent on task" or as a combination of number correct and time spent on task is also plausible. Unfortunately, the information of time spent on task is not stored in the currently available data and therefore can't be included in hypotheses. Interpreting only the currently variable number correct also neglects the fact, that different number of distraction messages have been presented. Only distracting message type 1 and 2 are displayed for all individuals reaching the critical items. Message type 3 depends on performance. Neither the purpose of the message type 1 and 2 nor the purpose of message type 3 is discussed in the current manual.

Under the achievement motive theory perspective the function of message type 1 and 2 is to prompt cognitions about one's endurance on the task, which corresponds with the goal of quantifying „endurance" by counting the number of correct discriminated figures. And at least the message type 1 expresses the „success" explicitly. Message type 2 which is also referring to the performance concerning attention is possibly experienced inconsistent by

some individuals. Why should individuals be reminded that terminating is possible, when „...
no longer able to concentrate on the task to work regular...“, when they are actually working
regular.

Under the same perspective the message type 3 has to be interpreted as experiencing
„failure“ and may not only prompt cognitions about endurance but also about one's perform-
ance concerning attention.

With this additional factor termination in an later phase of the test might not only repre-
sent more endurance, but also the existence of more attention (when somebody is working
with running power, without errors or performing irregularities.) Or it might indicate that
somebody is determined to – reasonable or not – continue even when he is no longer per-
forming well. Differing nine states of termination based on the number of given distracting
messages type 3 and the 4 phases will allow to include some effects of performance differ-
ences in the interpretation of endurance (Table 1). Important is that with this classification
individuals experiencing failure because of the feedback of performance irregularities can be
detected.

Table 1:
Different states of test termination - for figure discrimination

termination in phase	number of type 3 messages		
1	-	Extremely early termination	no success - feedback not experienced failure
2	0	Early termination without performance irregularities	no success - feedback not experienced failure
2	1+	Early termination with performance irregularities	no success - feedback experienced slight failure
3	0	Termination after distracting message type 1, without performance irregularities	success - feedback not experienced failure
3	1-2	Termination after distracting message type 1, with slight performance irregularities	success - feedback experienced slight failure
3	3-35	termination after distracting message type 1, with major performance irregularities	success - feedback experienced failure
4	0-1	termination after distracting message type 2, without performance irregularities	success-feedback not experienced failure, *inconsistent type 2 message!*
4	3-35	termination after distracting message type 2, with performance irregularities	success-feedback experienced failure consistent reminder type 2
4	36+	termination after distracting message type 2, with extreme performance irregularities	success-feedback experienced massive failure consistent reminder type 2

Although „endurance" is regarded as an important variable of the achievement motive by practitioners, studies dealing with endurance are rare. With some limitations it is possible to refer to the results of Feather (1961; 1963) and Schmalt (1999) described above, both dealing with endurance in the case of failure.

There are structural similarities between the task used in these studies and the FD-task but there are significant differences: Success on the FD-task does not rely on a testee merely solving it or not, but rather on the ongoing observation of the correct coding process. Also there is no objective criterion for failure on the FD-task. Additionally the task is rather easy (and boring) - the only challenge lies in „working longer than others" or „keeping the attention and working without errors".

Other results relevant for the content validity of the FD-task and its variables might be the results from two meta-analysis on the effects of goal setting on task performance (Locke, Shaw, Saari, & Latham, 1981; Mento, Steel, & Karren, 1987). It was found that goal difficulty and goal specificity/difficulty were strongly related to task performance and invested effort across a variety of tasks. According to Kleinbeck, Schmidt, & Carlsen (1985) the correlation is moderated through the achievement motive. Approach motivated individuals adapt their effort according to the goals and the difficulty, whereas avoidance motivated individuals did not modify their effort.

Referring to Feather (1961; 1963) and Schmalt (1999) results of approach motivated individuals are supposed to persist longer on the FD-task even when facing lots of distracting messages (experiencing failure) than avoidance motivated individuals. But with reference to (Kleinbeck et al., 1985) this persistence should be well adapted to be an indicator for approach motivation. Investing effort in an attention task, when one is currently no longer able to work with attention on the task, is bad adapted effort and is a sign of avoidance motivation.

Further insight into these tasks power to assess achievement motive can be gained by discussing aspects of motives activated by the material and motives present in the process of answering like Schmalt and Sokolowski (2000) did for TAT and MMG. When taking the TAT, the MMG or a questionnaire one has to imagine oneself in the shown or described situation in order to be stimulated by the material. Schmalt and Sokolowski (2000) rely on privileged processing of the pictorial material to claim the motive stimulating effect of TAT and MMG. The work style tasks are achievement tasks and should therefore activate the achievement motive, without any need to imagine oneself in the situation.

The question of whether this activation of achievement motive stays explicit or implicit becomes evident in the results of a study by (Hofmann & Kubinger, 2001). After working on a personality questionnaire or on the test battery „work style" subjects (comparable to the „research sample" described later) indicate which traits they think have been assessed by the tests they took by selecting appropriate variables from a list of 20 possibly assessed traits. The traits assessed with work style were identified by an average of 25% of the sample (maximum 41%), whereas the traits assessed with the questionnaire were identified by 67% to 93% of the sample. A comparison of the results of persons, who identified the intention of the tests with persons who did not identify the traits did not yield any systematic differences in the Work Style variables. It can therefore be concluded that for the majority of the subjects, the intention of assessing variables of achievement motive stays concealed. This indicates that achievement motive is activated implicitly. (Although not further discussed here, it must be said in this context, that concealing the intention of what is to be measured in an

assessment is problematic from the perspective of fairness and should be considered in further improvement of the test.)

Selecting the relevant semantic answer in the MMG requires more conscious systems, and is thus similar to the process in questionnaires. Certain parts of the process of answering in the work style tasks are claimed to be conscious decisions. Doing the CS-task one has to reflect on his/her performance and his/her ability in order to be able to come up with a prediction. And during the FD-task the decision to terminate or to continue the task has to be made. Unclear is whether the process leading up to these decisions (both the „prediction" and the „decision about continuing or not) remains unconscious and also whether this process is influenced by any situational factors.

Another similarity between MMG and questionnaires is that the answer is used „as it is", whereas the answers produced in TAT and Work Style tasks are further processed to form the various variables of interest. TAT results are scored using a classification system, Work Style performance results are used to compute the specially designed OPT-variables „frustration tolerance" and „aspiration level".

In the following several hypotheses concerning aspects of content validity of the CS and the FD task and variables are put forward and tested using a quasi experimental design with samples working under different motivational settings. Hypotheses are formulated on the following assumptions: One of the samples took the test voluntarily as part of a research project (research sample), the other sample took the test as part of the selection process for job or training (selection sample; see section „samples" for a detailed description). According to (Weinberger & McClelland, 1990) implicit motives are responsive to natural incentives in the situation. It's most likely that the implicit motive is activated in the exam-like situation of taking a computerised test. Also compare that some of the pictures used in MMG to stimulate the implicit achievement motive touch the situation „taking an exam" or „talk to director". Additionally the selection situation is full of social incentives (getting the job) who are said to stimulates the self-attributed part of the achievement motive (Weinberger & McClelland, 1990), but there are less social incentives stimulating the self attributed motives in the research situation. Therefore it is supposed that for individuals of the selection sample the achievement motive is more active than for the individuals of the research sample, which should result in different performance results.

To focus on the question of whether the achievement motive is activated in the CS-task the absolute performance results of the CS task are compared between the two samples.

First it is checked whether the two samples are equally skilled in coding symbols, by comparing the CS-trial 1 performance. It is the only trial where the individuals worked without task-specific manipulation, like „reflecting the performance" or „social comparison" and according to the manual CS trial 1 serves as warm-up trial not used for the computation of OPT-scores.

Hypothesis a

There are no mean performance differences in CS trial 1 between the research sample and the selection sample.

Mean performance over the trials 2 to 5 (as well as the prediction) is assumed to rise slightly (or to stay constant at least). The difficulty of the symbols is constant over the 5

trials and the coding intervals of 50 sec. are short, so there are no effects of tiredness assumed.

Hypothesis b

The mean performance (and the mean prediction) increases over the 5 CS trials for both samples.

The next hypothesis focus on the effect of the first performance feedback and performance prediction on the performance. Making a performance prediction after trial 1 requires at least some reflecting about the past performance and the intended effort for the next trial, both self-relevant cognitions, who are said to stimulate the self-attributed part of the achievement motive (Weinberger & McClelland, 1990). Together with the already very active implicit motive and the stimulation through social incentives the selection sample is expected to show better performance results than the research sample from trial 2 onwards. The socially more relevant situation of being in a selection process foster the effort that is put in the task.

Hypothesis c

In the CS trials 2 to 5 the selection sample is performing better than the research sample.

The same process should result in selection sample's higher prediction of trial 2 performance.

Hypothesis d

The selection sample shows more performance-prediction difference for trial 1 than the research sample.

For the effect of social comparison (after CS-trial 2 to CS-trial 5) on the prediction no hypothesis is currently formulated. There are no leads provided by literature for the kind of difference between the two samples when they are faced with the social comparison, observed results for the current samples are reported.

Although a quasi-experimental approach for validating the performance independent OPT-variables „frustration tolerance" and „aspiration level" is not enough this approach can give a first hint on the validity of these variables. The principle of OPT-variables is to create variables, that reflect differences in the structure of the performance as an indicator for personality traits, rather than reflecting level of performance as an indicator for ability. „Frustration tolerance" and „Aspiration level" have been created to reflect relevant aspects of the achievement motive. There is no reason why the two samples should differ regarding the underlying motives and there should be no differences between the two samples regarding these OPT-variables.

Hypothesis e

Regarding the OPT-variables „frustration tolerance" and „aspiration level" there are no differences between the two samples.

Deriving hypotheses to focus on the question of whether the achievement motive is activated in the FD-task provides some problems. The current structure of the task with feedbacks referring to endurance and attention makes it difficult to define whether the performance in this task reflects endurance as a personality construct or rather reflects the ability of keeping the attention on the task.

Under the assumption that working on the task and terminating the test is mainly influenced by the achievement motive the selection sample is supposed to show more performance (higher number of correct discriminated figures) than the research sample. Selection sample' s achievement motive is already activated through situational cues and social incentives, additionally there are distracting messages to stimulate selection sample's cognitions about their endurance (and attention) on the task.

Hypothesis f

The number of correct discriminated figures is higher for the selection sample, than for the research sample.

Under the assumption, that working on the task and termination of the test is influenced not only by the achievement motive but also by cognitions concerning the likely tested ability – keeping attention on the task – the information about losing the attention is supposed to influence the samples in different ways. For the selection sample the importance of losing their attention is defined through the situation – they are applying for jobs where attention is important and they have to prove that they can work with staying power. So when deciding termination or not they carefully have to balance between showing endurance and performing well. For the research sample neither the importance of effort nor the relevance of attention is defined through the situation. For them the balance between showing endurance and performing well is not so important. To quantify this „balance" the variable number of correct discriminated figures is not suitable. One has to use a variable, that integrates information about the phase (number of discriminated symbols) and the conditions (number of type 3 messages) when the individuals decides to terminate the test.

A sensible point for terminating the test, specially when one has already experienced performance irregularities is in phase 3, after the type 1 message. The message gives feedback about the success of working till here, and after some feedbacks about losing attention, terminating the test is a sensible solution for somebody interested in balancing showing endurance and performing well. For the selection sample this balance is of more importance than for the research sample, it is therefore expected that more individuals from the selection sample terminate the test in this state.

Hypothesis g

More selection sample subjects than research sample subjects terminate the FD-task in state 33-35 (Termination after distracting message type 1, with major performance irregularities).

Materials and Method

Samples

All Work Style data sets available for analysis were aggregated to form 2 samples: Sample a – 'selection sample' of young people (n=408; age: 17-32, mean: 29.6) taking the test as an application for different kinds of training. One part of this sample took part in the selection and assessment for training as air traffic control staff (AustroControl; Vienna, Austria); the test battery 'Work Style' was part of the multi-level assessment. The data was kindly placed at this study's disposal by Hoffmann and Schrott (in prep.) who are currently evaluating the air traffic control staff assessment and selection. The other part of this sample was placed at this study's disposal by the 'Deutsche Bundeswehr' (Deutsche Bundeswehr, 2000). A second sample (n=757; age: 17-60, mean: 26.6; sample b-'research-sample') consisted of persons who had volunteered for various research projects (University of Vienna, Austria) as a part of their training (students studying psychology) or just out of interest. The results of the testing did not have any consequence for them and were kept anonymous.

To conduct analysis for the CS-task, sub-samples with matched age structure were taken at random out of sample a (n=92; 24 female, 68 male, age: 18-28, mean: 20.3) and sample b (n=92, 16 male, 76 female, mean=22.58). When comparing absolute performance results especially with regard to speed as the main component, the influence of the age must be considered in the analysis. The sub-samples are similar concerning age, and the individuals are assumed, in principal, to be equally skilful in handling the mouse to work through the test. As the factor sex is linked with the situation and could not be balanced, it will be considered as a covariate.

Material

People from all samples took the test 'Work Style' as part of a larger computerised test-battery together with other tests and questionnaires, not of relevance in the context of this study.

Method

As this study has not been planed as such, there were certain limitations regarding the planing of the method. The fact that the available samples worked under different settings, which allowed for an excellent comparison between two different motivational settings, was significant for putting up hypotheses to contribute to the content validity of the two tasks. The situation „real selection" cannot be investigated experimentally, but the comparison of the results from real situations with results from research context is of practical relevance for the validity of variables. Naturally there are certain situational influences that are relevant when taking a test, for example, more or less stimulating situations to achieve good results. The motivation to gain good results is likely to be higher in a personnel-selection situation than in a research context: People from sample a took the test 'Work Style' together with other tests during the selection process. This sample's desire to obtain good results is consid-

ered as high (internal and external factors). Sample b's desire to obtain good results is claimed to be driven by internal factors only, as the result of the test had no external consequences for them.

Results

Subtest 2 – Coding Symbols

The selection sample makes an average of up to 6 more correct symbols (*Notes: sub-sample selection: n=92; 24 female, 68 male, age: M=20.3; SD=1.98; sub-sample research: n=92; 16 male, 76 female, age: M=22.58; SD=1.58.; trial 1=first performance, first prediction.*

Figure 3The mean differences are significant (multivariate analysis of variance, 5 trials as 5 variables, with sex as covariate: Wilks-Lambda=0.9; F=3.944, df(hypothesis)=5; df(error)=177; p=0.002; Box-M-test=13.691; F=0.886; df=15; p=0.580). Single comparison of means indicates differences for trial 3, 4 and 5, with higher coding performance for the selection sample. Sex as a covariate had no significant influence on the absolute coding performance. These results reveal that the two samples are equally skilled in coding symbols (hypothesis a). They also show that the selection sample shows better performance with ongoing trials, tough not as supposed in hypothesis c from trial 2 onwards.

As displayed in Figure 3 the mean performance over the 5 trials increases for both samples (hypothesis b), but more for the selection sample (univariate repeated measurement design with a covariate, within-subject factor: 5 trials 5 time points; between-subject factor: sample;: Wilks-Lambda=0.152; F=249.173; df(hypothesis)=4; df(error)=179; p=0.000; Box-M-test= 13.691 ; F= 0.886; df1= 15; p= 0.580), but there is also effect between the two samples (Wilks-Lambda=0.887; F=5.721; df(hypothesis)=4; df(error)=179; p=0.000).

Data presented in Figure 3 shows that the mean prediction exceeds the mean performance for both samples over the 5 trials, with the selection sample making even higher predictions than the research sample. To test the effects of the situation on the performance and the prediction over the five trials, a multivariate repeated measurement design with a covariate was defined (within-subject factor: 2 measures: performance, prediction; 5 trials; between-subject factor: sample; covariate: sex). The significant Box's M indicates that the assumption of the same variance-covariance matrices across the cells is violated. It was than decided to use a univariate repeated measurement design with a covariate (within-subject factor: performance-prediction pair (pp-pair) 2 time points; between-subject factor: sample; covariate: sex) instead. It should be mentioned here that the prediction is defined as a repeated measure. Results for each trial are displayed in Table 2.

For the trials 1, 4 and 5 Box's M is insignificant, therefore the significant F-Tests for the effects are interpreted: For trial 1 the significant main effect pp-pair and the insignificant pp-pair*sex interaction indicates that observed differences in the performance-prediction relation are not related to sex. Therefore, the significant result for the pp-pair*sample interaction indicates that there are differences in the performance-prediction relation between the two situations: the selection sample makes the more optimistic predictions, at least, when they are setting their aspiration level by internal standards only (hypothesis d).
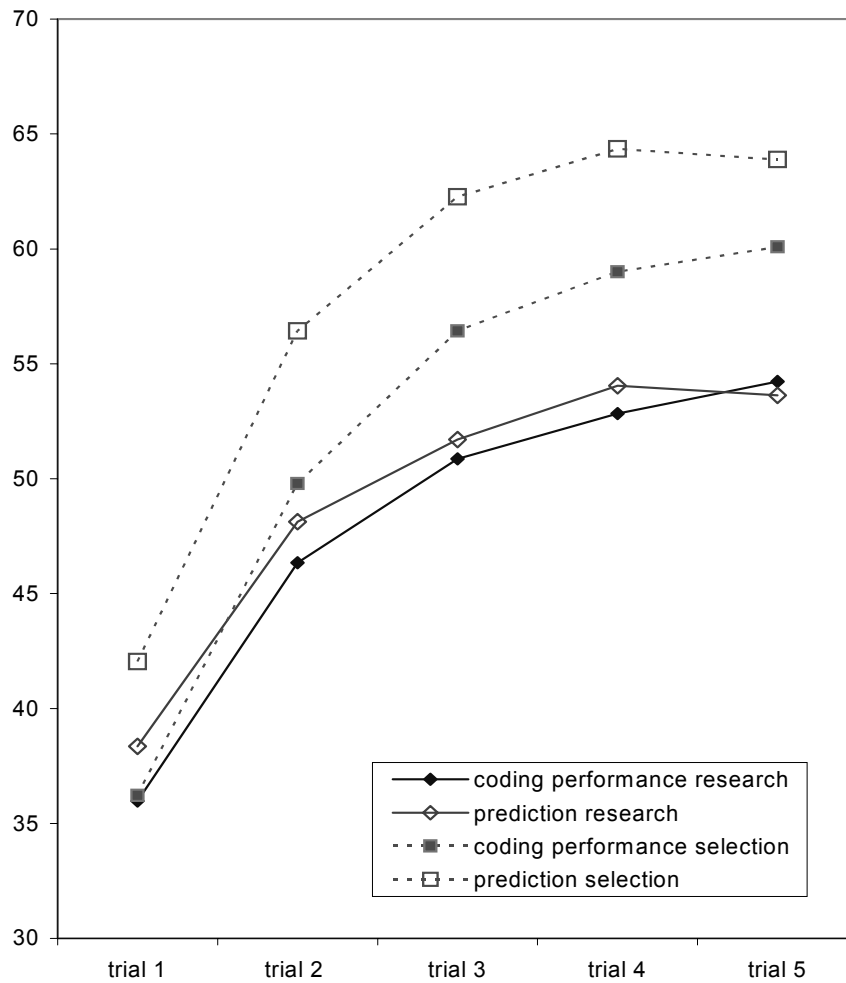
Notes: sub-sample selection: n=92; 24 female, 68 male, age: M=20.3; SD=1.98; sub-sample research: n=92; 16 male, 76 female, age: M=22.58; SD=1.58.; trial 1=first performance, first prediction.

Figure 3:
Mean coding performance and mean prediction for the selection and the research
sample per trial

Table 2:
Testing the effects of situation on the performance and the prediction over the five trials:
univariate repeated measurement design with covariate

	Box-M-Test		Effect	Wilks-Lambda	F	
trial 1	Box's M	4.805				s.
			PP-PAIR	0.754	58.981	
	F	1.583	PP-PAIR * SEX	0.985	2.685	
			PP-PAIR * SAMPLE	0.960	7.528	s.
trial 2	Box's M	15.297				s.
			PP-PAIR	0.842	33.870	
	F	5.038	s. PP-PAIR * SEX	0.978	4.057	s.
			PP-PAIR * SAMPLE	0.975	4.693	s.
trial 3	Box's M	8.363				s.
			PP-PAIR	0.879	24.924	
	F	2.754	s. PP-PAIR * SEX	0.982	3.381	
			PP-PAIR * SAMPLE	0.968	6.032	s.
trial 4	Box's M	1.700				s.
			PP-PAIR	0.860	29.577	
	F	.560	PP-PAIR * SEX	0.965	6.505	s.
			PP-PAIR * SAMPLE	0.988	2.188	
trial 5	Box's M	4.488				s.
			PP-PAIR	0.860	29.577	
	F	1.478	PP-PAIR * SEX	0.965	6.505	s.
			PP-PAIR * SAMPLE	0.988	2.188	

Notes: within-subject factor: performance-prediction pair (PP-PAIR), prediction is defined as repeated measure; between-subject factor: research sample-selection sample (SAMPLE); covariate: sex (SEX)
s.: p<.05

Due to the fact that the first negative social comparison is made after trial 2, the results of the statistical test for performance-prediction difference in trial 2 and trial 3 are relevant. The heterogeneous variance-covariance-matrices might be an indicator that the 'treatment effect' influenced the samples differently, and are not interpreted here. For trials 4 and 5 the significant pp-pair*sex and insignificant pp-pair*sample interaction indicates sex related differences rather than situational differences.

Making an optimistic prediction is one part of the story, being able to reach this prediction is another. Although the current study has no literature focusing on „reaching a prediction" at disposal the prediction-performance-relationship is displayed (Figure 4) and analysed for the two samples. Pairs of predictions and the following performance (PrPe-pairs, e.g. prediction after trial 1 and performance of trial 2 is written as trial 1/2) were built and analysed with the same approach as the performance-prediction pairs.
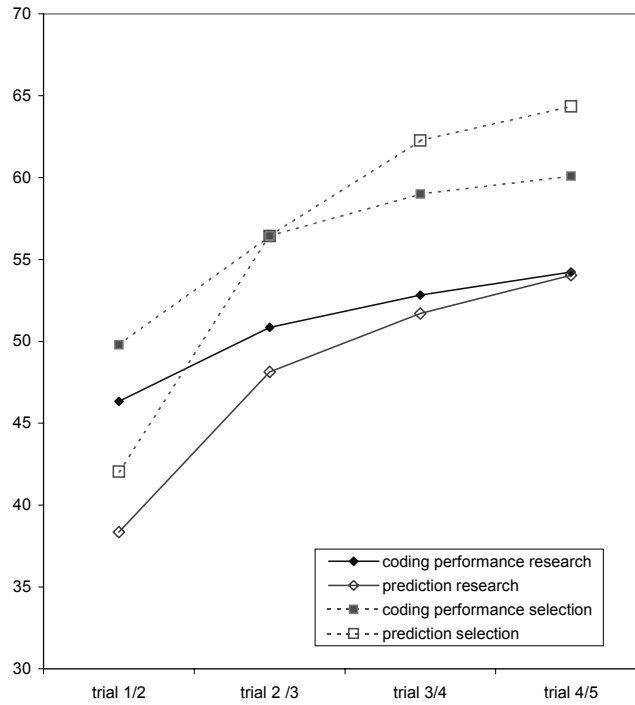
Notes: sub-sample selection: n=92; 24 female, 68 male, age: M=20.3; SD=1.98; sub-sample research: n=92; 16 male, 76 female, age: M=22.58; SD=1.58.; trial1/2: first prediction; second performance.

Figure 4:
Mean prediction and the following mean performance for the selection and the
research sample

Results for the main effect in trial 1/2 and 4/5 indicates that prediction and following performance differ significantly. For trial 1/2 there are no situational influences or sex influences - performance exceeds the prediction. For trial 4/5 sex has an effect, but not the situation: Males' coding performance stays behind their predictions. What is of interest is the situation for trial 2/3 and 3/4. For trial 2/3 there is no main effect, and no interaction with the situation, but sex does have an influence. Females still exceed their prediction, males do not. For trial 3/4 there is no main effect and no effect of sex, but a significant interaction with the situation. Individuals working under a selection situation remain behind their prediction.

The mean differences regarding the objective variables 'aspiration level' and 'frustration tolerance' (hypothesis e) are insignificant (multivariate analysis of variance, with sex as covariate: Wilks-Lambda=0.990; F=0.928, df(hypothesis)=2; df(error)=180; p=0.39; Box-M-test=34.894; F=11.478; df1=3; p=0.000), but the significant Box-M test might invalidate this result. However, a univariate comparison showed no difference between the two samples (Table 4).

Table 3:
Testing the effects of situation on the prediction and following performance for the 4
prediction-performance pairs: univariate repeated measurement design with covariate

	Box-M-Test		Effect	Wilks-Lambda	F	
trial 1/2	Box's M	3.579	PrPe-pair	.785	49.616	s.
	F	1.179	PP-PAIR * SEX	.999	.185	
			PP-PAIR * SAMPLE	1.000	.008	
trial 2/3	Box's M	5.125	PrPe –PAIR	.996	.684	
	F	1.688	PP-PAIR * SEX	.964	6.842	s.
			PP-PAIR * SAMPLE	1.000	.057	
trial 3/4	Box's M	5.125	PrPe –PAIR	.983	3.043	
	F	1.688	PP-PAIR * SEX	.995	.890	
			PP-PAIR * SAMPLE	.976	4.392	s.
trial 4/5	Box's M	.208	PrPe –PAIR	.920	15.760	s.
	F	.069	PP-PAIR * SEX	.958	7.916	s.
			PP-PAIR * SAMPLE	.995	.949	

Notes: within-subject factor: prediction-performance pair (PrPe-PAIR), performance is defined as repeated measure; between-subject factor: research sample-selection sample (SAMPLE); covariate: sex (SEX)
s.: p<.05

Table 4:
Descriptive statistics and univariate tests - for the objective variables of coding symbols

variable	mean	stdev	median	percentile 25/75	percentile 10/90	min/max
aspiration level						
sample b1 - research	-.12	.55	-.17	-.27 / -0.06	-.41 /.00	-.85 / 3.50
sample a1 - selection	-.12	.49	-.15	-.25/-0.02	-.44/.06	-.81 / 4
t-test (homogeneous variances: F=0.028; p=0.867): T=0.003; df=182; p= 0.998 n.s.						
frustration tolerance						
sample b1 - research	.14	.30	.11	.005 / .24	-.14 / .54	-.91 / .88
sample a1 - selection	.21	.50	.14	.00 / .22	-.03/.38	-.29 / 4
t-test (homogeneous variances: F=0.448; p=0.504) T=-1.108; df=182; p=0.269 n.s.						

Notes: sub-sample selection: n=92; 24 female, 68 male, age: M=20.3; SD=1.98; sub-sample research: n=92; 16 male, 76 female, age: M=22.58; SD=1.58.
s.: p<.05

The mean of -.12 for the aspiration level indicates that in general the performance exceeds the prediction by about 12%. The mean of .14 (.21) for frustration tolerance indicates that the fifth prediction is about 14% (21%) higher than the second prediction.

Subtest 3 - Figure discrimination

The results of figure discrimination are based on the complete research and selection sample due to the fact that speed is not the essential variable for this endurance task.

Comparison of the number of correct discriminated symbols between the two samples yields no difference in discrimination performance (t-test, heterogeneous variances: F=9.418; p=0.002; T=-1.585; df=941.654; p= 0.113). According to hypothesis f this result indicates, that working on the FD-task and terminating the task is not only influenced by the achievement motive.

To investigate the alternative assumption (hypothesis g), that working on the task and termination of the test is influenced not only by the achievement motive but also by cognitions concerning the likely tested ability – keeping attention on the task – the percentage of persons terminating the test in one of the important states are compared (Table 5).

The percentage of subjects terminating in state 3 3-35 (termination after distracting message 1, with major performance irregularities) is higher in the selection sample. Additionally, the percentage of subjects termination in state 1- and 20 (extreme early termination and early termination without performance irregularities) is lower (Table 5).

Table 5:

Frequencies in percent for different states of test termination – for figure discrimination

	% subjects terminating the test in ...								
	state 1-	state 2 0	state 2 1+	state 3 0	state 3 1-2	state 3 3-35	state 4 0-1	state 4 3-35	state 4 36+
selection-sample (n=406)	6.7%	2.5%	3.9%	3.9%	17.2%	26.1%	34.0%	3.0%	2.7%
research-sample (n=748)	9.1%	6.7%	2.8%	4.9%	16.0%	17.4%	36.8%	3.5%	2.8%
$\chi^2 = 23.384$ s. (minimum expected frequency = 11.26); df = 8									

Notes:
state 1-: extremely early termination
state 2 0: early termination, without performance irregularities
state 2 1+: early termination, with performance irregularities
state3 0: termination after distracting message type 1, without performance irregularities
state 3 1-2: termination after distracting message type 1, with slight performance irregularities
state 3 3-35: termination after distracting message type 1, with major performance irregularities
state 4 0-1: termination after distracting message type 2, without performance irregularities
state 4 3-35: termination after distracting message type 2, with performance irregularities
state 4 36+: termination after distracting message type 2, with extreme performance irregularities
s.: p<.05

Discussion

The study reported here has been designed so as to increase the understanding of a new technique for assessing achievement motive. The idea behind this new technique was to apply Cattell's objective personality testing approach in the field of achievement motive. The OPT approach relies on observing behaviour rather than asking questions about behaviour. The subtests Coding Symbols (CS) and Figure Discriminations (FD) from the test „Work Style" have been designed so as to observe behaviour in an achievement situation. Within CS-task performance- feedback and goal setting is used to observe the aspiration level of the testee, and a social comparison of the testee's performance is used to induce frustration. In the FD-task, feedback is given to stimulate the reflections upon one's own performance. The OPT-variables derived claim to assess „aspiration level", „frustration tolerance" and „endurance" in order to indicate the underlying achievement motive. These variables do not reflect differences in performance. The absolute performance results are standardised individually.

Validating OPTs requires us to first ensure the content validity of the task used to derive OPT-variables. A quasi experimental design using samples working under different motivational settings was used to formulate hypotheses focusing on the content validity of these two tasks. The idea was to find out whether the achievement motive had been stimulated by the tasks and the key approach was to compare the absolute performance results between the two samples. The assumption about performance differences between the two samples is driven by the idea that the implicit part of the achievement motive is activated in both the research situation and the selection situation. The self-attributed part of the motive is activated more strongly in the selection situation, as there are more social incentives, which are said to stimulate this part of the motive (Weinberger & McClelland, 1990).

The precondition for comparing performance differences - so that achievement motive can be said to be activated by the task - is to have two equally skilled samples. With no performance differences observed in CS-trial 1, this was fulfilled. It was hypothesised that the task and the feedback stimulates achievement motive for both samples which should result in increasing performance. The current results show a rise in the mean performance over the trials 2 to 5 for both samples. The more specific hypothesis about the higher performance level of the selection sample in comparison to that of the research sample was also verified. The performance level of the selection sample is higher than that of the research sample - but only for the trials 3 to 5, not for trial 2. The lack of effect regarding performance in trial 2 might be compensated by the selection sample's more optimistic predictions after trial 1. At least this value indicates that the achievement motive is activated more strongly for the selection sample.

No hypothesis was formulated for the effect of social comparison (after CS-trial 2 to CS-trial 5) on the prediction of the testee. Results observed in this study are heterogeneous between the trials and indicate certain sex related differences rather than situational differences, which should be investigated separately. Another observation is that the prediction and the following performance differ significantly. Both samples' performance levels exceed their first prediction (after trial 1) in trial 2, although their first prediction already includes an anticipated increase in performance. However the selection sample seems to continue to anticipate a performance increase with ongoing trials: In trial 4 only these individuals remain behind the prediction that they gave after trial 3. For all other prediction performance pairs

the current results are heterogeneous and again indicate sex related differences rather than situational differences. In general, these findings show that the CS-task is suitable to activate the achievement motive. Therefore, it can be assumed that the content validity for the OPT-variables „frustration tolerance" and „aspiration level" is given.

Regarding other aspects of validity, such as construct validity, the quasi-experimental approach can not provide any results. However, as assumed, the two samples do not differ regarding these variables. This might be a first indicator, that they reflect achievement motive rather than resulting achievement motivation.

Results for the FD-task indicate that the variable „endurance" does not represent endurance as an indicator of the resulting achievement motivation. This is due to the fact that the structure of the task not only shows a lot of similarity with tasks used for assessing attention, but also stimulates the reflections upon one's own attention when working on the task. Using only the information of the number of correctly discriminated figures does not represent the motivational situation, which stimulated the actual termination of the test. When integrating the information about the attention performance, as suggested above, there is a difference between the two samples observable. The intention of the authors of „Work Style" was to „....define a task, which is easy to reduce the influence of „ability", but allows the person to show „endurance" or „effort". But showing effort and endurance implies that there are difficulties that have to be overcome and that the individual has to balance effort and endurance with the subjective probability of success or failure on the task. The difficulty of the FD-task is „keeping one's attention on the task", not the task itself. The subjective probability of keeping one's attention on the task is manipulated only for some individuals through the distracting messages typ 3. In the literature cited in the task-description above, the subjective probability was manipulated for all persons through the instruction. Constructing a powerful endurance and effort task based on the paradigm of shifting subjective probability for success and failure requires the possibility to manipulate success and failure independently of the ability on a task, but similarly for all subjects taking the task.

In summary it can be concluded that the CS-task is currently suitable for evoking achievement motive, which is the precondition for the construction of content valid OPT-variables. The first allusion to the construct validity of the OPT-variables formed from the CS-task is also positive. The strategy of the CS-task of using a goal setting task to draw conclusions on the underlying achievement motive looks promising. Further research should focus on characterising these variables in the context of traditional achievement motive measures as well as personality measures to gain more information on construct validity as well as convergent and discriminant validity.

For the FD-task, differences between the two samples were observed regarding the balance between the performance of keeping one's attention on the task and the effort and endurance spent on the task. This indicates that the task evokes achievement motive, which is the precondition for the construction of valid OPT-variables. The variable „endurance" defined in the current version of the test is not valid for this purpose, as it is confounded with the ability to keep one's attention on the task. Further development of the FD-task is necessary to increase the content validity in assessing „endurance" and „effort".

References

1. Atkinson, J. W. (1957). Motivational determinants of risk-taking behaviour. Psychological Review, 64, 359-372.
2. Atkinson, J. W. (1964). An introduction to motivation. Princeton, N.J: van Nostrand.
3. Cattell, R. B. (1955). The O-A (Objective-Analytic) Personality Factor Batteries: Adult and Child Forms. Champaign: IPAT.
4. Cattell, R. B. (1958). What is „objective" in „objective" personality tests. Journal of Counseling Psychology, 5, 285-289.
5. Cattell, R. B., & Warburton, F. W. (1967). Objective personality and motivation tests. A theoretical introduction and practical compendium. Chicago: University of Chicago Press.
6. Feather, N. T. (1961). The relation of persistence at a task to expectation of success and achievement related motives. Journal of Abnormal and Social Psychology, 63, 552-561.
7. Feather, N. T. (1963). Persistence at a difficult task with alternative task of intermediate difficulty. Journal of Abnormal and Social Psychology, 66(604-609).
8. Heckhausen, H. (1989). Motivation und Handeln. [Motivation and Action]. Berlin: Springer.
9. Heckhausen, H., Schmalt, H.-D., & Schneider, K. (1985). Achievement motivation in perspective. New York: Academic Press.
10. Hofmann, K., & Kubinger, K. D. (2001). Herkömmliche Persönlichkeitsfragebogen und Objektive Persönlichkeitstests im „Wettstreit" um (Un-)Verfälschbarkeit (Conventional personality questionnaires and objective personality tests competing for (non)falsifiability). Report-Psychologie, 26(5-6), 298-304.
11. Kleinbeck, U., Schmidt, K.-H., & Carlsen, H. (1985). Veränderungen von Zielsetzungswirkungen auf die Leistung durch leistungsthematische Einflussfaktoren. Zeitschrift für Experimentelle und Angewandte Psychologie., 32, 263-280.
12. Kubinger, K. D. (2002). Personality questionnaires: Some critical points of view [Special Issue]. Psychologische Beiträge, 44(1).
13. Kubinger, K. D., & Ebenhöh, J. (1996). Arbeitshaltungen. Kurze Testbatterie: Anspruchsniveau, Frustrationstoleranz, Leistungsmotivation, Reflexivität/Impulsivität. [Work Styles. Short test battery: Aspiration level, resistance to frustration, achievement motivation, reflexivity/impulsivity. English Version: In: The Vienna test system. Mödling: Dr.G.Schuhfried.GmbH.]. Frankfurt: Swets.
14. Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969 -1980. Psychological Bulletin, 90, 125-152.
15. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). The achievement motive. New York: Appleton-Century-Crofts.
16. McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? Psychological Review(96), 690-702.
17. Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966-1984. Organizational Behaviour and Human Decision Processes, 39, 52-83.
18. Murray, H. A. (1942). Thematic apperception test. Cambridge, Mass.: Harvard University Press.
19. Schmalt, H.-D. (1999). Assessing the achievement motive using the grid technique. Journal of Research in Personality, 33, 109-130.
20. Schmalt, H.-D., & Sokolowski, K. (2000). Zum gegenwärtigen Stand der Motivdiagnostik. [Current positions in assessing motives]. Diagnostica, 46(3), 115-123.

21. Schmalt, H.-D., Sokolowski, K., & Langens, T. (2000). Das Multi-Motiv-Gitter zur Erfassung von Anschluss, Leistung und Macht - MMG. [The multi-motive-grid for the assessment of achievement-, affiliation- and power motive - MMG]. Frankfurt: Swets.

22. Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. Psychological Bulletin, 112, 140-154.

23. Weinberger, J., & McClelland, D. C. (1990). Cognitive versus traditional Motivational Models. Irreconcilable or Complementary? In E. T. Higgins & R. M. Sorrentino (Eds.), Handbook of motivation and cognition (Vol. 2, pp. 562-597). New York: Guilford Press.