Think-Aloud Approaches to Cognitive Assessment and the Articulated Thoughts in Simulated Situations Paradigm

Gerald C. Davison, Ralph S. Vogel, and Sandra G. Coffman
University of Southern California

In addition to widely used endorsement methods, one way to get at people's thoughts is to have them verbalize while engaged in a task or situation. The articulated thoughts in simulated situations (ATSS) paradigm is a think-aloud approach to cognitive assessment that has several advantages: an unstructured production response format, on-line rather than retrospective assessment, situational specificity and control, and flexibility of situation and cognitions. The authors review experiments that have examined articulated thoughts in clinically relevant contexts. ATSS does have certain limitations and further research into its psychometric properties is needed, but it seems promising as a versatile and adaptable method of cognitive assessment, especially when little is known of the cognitive terrain of interest.

What is on your mind? What were you thinking when you encountered that situation you have been dreading? How did you arrive at that conclusion? These are but a few of the questions that cognitive and clinical psychologists have put to participants and patients for many years and in many different contexts. The answers to such queries inform psychologists’ speculations about the role of thought in emotion and behavior.

The core assumption in cognitive–behavioral theory, research, and practice is that affective and behavioral responses are mediated by thought processes, both conscious and unconscious (Bock, 1967; Ellis, 1962). Operating within this paradigm, researchers have focused in recent years on the development of cognitive assessment methods but with little consensus in the approach to the measurement of cognition. As a result, vastly different modes of cognitive assessment have proliferated, without adequate attention to validity issues. Indeed, measures of the same constructs often do not correlate highly (Chamberlain & Haaga, in press; Clark, 1988).

In general, cognitive scientists as well as some clinical scientists have developed performance measures to infer cognitive processes. Experimental situations are created to make inferences about underlying cognitive processes. An example of this is a classic study by Bransford and Johnson (1973), in which all participants read the same description of a man getting ready in the morning to leave his house—he has breakfast, reads the newspaper, and so forth. Some participants are then told he is a stockbroker, others that he is unemployed. In the former condition, participants “remember” afterward that he was reading the financial pages; in the latter, the want ads. In fact, the story did not specify what parts of the newspaper he was reading. These findings were used to infer the existence of a schema that distorted the participants’ memories. In contrast to this performance-based approach, clinical researchers often attempt to assess cognition more directly, using self-reports in interviews and on questionnaires that require participants to introspect on both current states (e.g., “What do you think of that?”) and on how they retrospectively view their thoughts and feelings over a long period of time (e.g., “When you are being watched, what thoughts go through your mind?”).

One mode of cognitive assessment, the think-aloud approach, is viewed as particularly useful in understanding the products as well as the processes of cognition. Because think-aloud methods assess cognitions concurrently with their occurrence, they may be better suited to tapping actual thought content than other modes are. In a standard think-aloud method, researchers have participants verbalize cognitions while performing some task, and responses are then recorded for subsequent evaluation. For example, early problem-solving researchers had participants verbalize their thoughts while solving complex problems (Duncan, 1945), and later researchers had participants think aloud while trying to perceive correctly an ambiguous picture that was slowly being brought into focus (Davison, 1964; for a general review, see Ericsson & Simon, 1993). Reminiscent of Piaget’s (1954) pioneering work with children, in which he generated hypotheses about their thinking based on their natural tendencies to talk out loud to themselves while solving problems, think-aloud methods have also been used in education research for...
the assessment of cognitive processes like attention and reading comprehension as well as for intervention to improve comprehension, for example, using think-aloud methods to teach students to attend to and appraise relevant cues (Montague, 1993; Randall, Fairbanks, & Kennedy, 1986).

Several thought-sampling approaches began to be pursued by clinical researchers in the late 1970s, and they yielded useful information on the relationships among thought, emotion, and behavior. For example, in a study of patients who have had cardiac catheterization, Kendall et al. (1979) found that poor adjustment correlated with endorsements of negative self-statements on a self-report inventory. Using a more open-ended format, Sutton-Simon and Goldfried (1979) had participants with acrophobia write down what they believed their thoughts and feelings would be in a variety of height situations; fear of heights correlated with negative self-statements. What has come to be called thought listing was used by Cacioppo, Glass, and Merluzzi (1979), who had male participants write down their thoughts immediately prior to interacting with a female confederate. Those who scored high on a questionnaire of social anxiety generated more negative self-statements than did those with low scores. Finally, in an approach closest to our research, Hurlburt (1979) and Klinger (1978) innovated with the assessment of thoughts as they occurred in naturalistic situations (see Hurlburt, 1997). As they went about their daily activities, Hurlburt's participants carried with them a small random-interval tone generator that cued them to record in a small booklet their thoughts or feelings as they were aware of them at the moment. One of his many findings was that contrary to the stereotype, male participants reported very low frequencies of thoughts that were sexual in nature.

As our work on articulated thoughts began in 1978, we attempted to minimize or eliminate some of the limitations that we believed were inherent in these thought-sampling procedures, most especially the restrictions placed on participants' reports by requiring them to endorse experimenter-defined categories of thoughts, the practical limits placed on situations in which their thoughts might be sampled, and the lack of control over situations of possible theoretical interest—a particular problem when an event such as being criticized by a certain person might be infrequent enough as to be missed by in vivo thought sampling (Davison, Robins, & Johnson, 1983; Merluzzi, Rady, & Glass, 1981).

The ATSS Paradigm

Our own efforts over the past 20 years have entailed the application of a paradigm dubbed articulated thoughts in simulated situations (ATSS; Davison et al., 1983). ATSS is termed a paradigm because it entails a general approach to think-aloud cognitive assessment that is not wedded to specific procedures or technologies. For example, we and others have restricted ourselves so far to presenting simulated situations on audiotape rather than videotape. It cannot be assumed that better results would be obtained with videotape because allowing participants to create their own images of simulated situations presented aurally may permit imaginal scenarios of greater personal relevance and meaningfulness (as those people who grew up listening to serials on radio can attest to). In any event, the use of videotaped situations would be within the boundary conditions of the ATSS paradigm if the experimental situation is structured as we have done. When we refer to ATSS in this article, it is to the procedure about to be described that we are referring. The following exemplifies instructions given to participants:

In this study we are interested in the kinds of thoughts people have when they are in certain situations. Often, when people are going about their daily affairs, interacting with others and so forth, they have a kind of internal monologue going through their heads, a constant stream of thoughts or feelings which reflect their reactions to something which is happening.

What we'd like you to do is to play a part in a couple of situations we have taped. Your part will involve listening to situations and tuning in to what is running through your mind, and then saying these thoughts out loud. The tapes are divided into seven segments.

At the end of each segment, there will be a tone, followed by a pause of thirty seconds, during which time we would like you to say out loud whatever is going through your mind. Say as much as you can until you hear another tone. Of course, there are no right or wrong answers, so please just say whatever comes to mind, without judging whether it seems appropriate or not. The more you can tell us the better.

Try to imagine as clearly as you can that it is really you in the situation right now. Note that your task is not to speak back to any one of the voices on the tape, as though you were having a conversation with one of them. Rather, you should tune in to your own thoughts and say them out loud. The microphones in front of you will enable us to tape your comments.

Because thinking aloud in this manner is an unusual task, we generally precede the experimental tapes with a practice tape to orient participants to the procedure as well as to allow screening for compliance with the task. Virtually no participants are non-responsive, and the vast majority think aloud in the desired way after the instructions are given and the practice tape played, during which time the experimenter freely coaches and exhorts participants to verbalize their thoughts in as much detail as they can. Participants sometimes use the subjunctive mode, for example, "If someone said that to me, I would really think I was a jerk." We consider such an outcome undesirable because it distances the participant from the immediacy of the simulated situation. The experimenter tries to discourage this from happening with appropriate prompts such as "Rather than say what you would do if you were in the situation, try to believe that you actually are in the situation and then say whatever is going through your mind." After completion of the practice tape, our experience has been that further prompts very seldom need to be given to participants to keep them on task.

Each stimulus tape runs from 2 to 3 min, with each segment ranging from 10 to 15 s. In our research, we have used from five to eight segments per scenario. However, these procedural details are not cast in stone. We and others have used a variety of stimulus tapes, depending on the cognitive constructs of interest (as outlined below). For example, the "social criticism" tape was designed to elicit cognitions related to social anxiety. This tape begins by saying that the participant is at a social function and overhears two acquaintances talking about him or her. As the situation unfolds, these two acquaintances repeatedly berate the participant's behavior, manner of dress, and personality. The "garage" tape was designed to elicit angry, hostile, or aggressive cognitions. This tape involves being made to wait...
Advantageous Features of ATSS

In comparison with other cognitive assessment approaches, the ATSS paradigm has a number of desirable features that information-processing theory and research suggest can facilitate the accurate assessment of people's thoughts (Davison et al., 1983; Merluzzi et al., 1981). These characteristics concern the extent to which participants are constrained in the data they provide, the degree to which participants' reports are retrospective, experimenter control of situations in which cognitions of interest are likely to occur, and the flexibility of the approach in studying a broad range of cognitions.

Unstructured Response Format

Unlike the most prevalent approach to cognitive assessment, paper-and-pencil endorsement methods, think-aloud approaches such as ATSS allow participants great latitude in the data they provide. That is, with endorsement methods (e.g., the Automatic Thoughts Questionnaire [Hollon & Kendall, 1980]), participants select from experimenter-provided options that most closely reflect their thinking. The information obtainable from respondents is constrained by a priori decisions made before participants have an opportunity to express themselves. The provision of predetermined alternatives also provides prompts to participants for how they might respond; this is especially the case in a forced-choice format, where participants have to choose from among a limited set of predetermined options. In contrast, ATSS participants are asked and are helped to report all cognitions (i.e., open-ended responding) rather than being limited to experimenter-selected alternatives that may not be representative of their actual thoughts or feelings. The use of unstructured responses increases the potential for obtaining a rich sample of actual cognitions of interest, which can turn out to include things that were not foreseen by the experimenter. Given how little is known at this time about cognition and psychological disorder, it seems preferable to construct a paradigm that gives participants as much leeway as possible in the kind of data they may provide while still creating enough structure to make sense out of the data. Constraints on ATSS data are imposed later through strategies to analyze the content, according to experimenter interest, and a virtually unlimited number of different coding schemes can be used on the same data set.

On-Line Assessment

ATSS uses a near-concurrent approach to cognitive assessment. Thinking aloud that immediately follows each brief segment taps cognitions as close to on-line as possible, without interfering with the listening task (and thus minimizing reactivity). A division of ATSS stimulus tapes into short segments facilitates retention of small chunks of information in short-term memory during each 30-s think-aloud period. Reduction of the participant's reliance on long-term memory and on retrospective reporting of generalized patterns of thinking over time minimizes, we believe, censoring and distortions (e.g., self-presentation bias, confabulation) — a common problem with paper-and-pencil measures, which require the participant to reflect back over a host of situations and then draw a general conclusion about his or her characteristic ways of thinking. The on-line character of ATSS and the ease with which participants can meet its requirements also make it suitable for concurrent psychophysiological measurement. For example, in a dissertation currently in preparation (Vogel, 1997), ATSS is used to assess self-efficacy concurrently with assessment of cardiovascular function, including tracking-cuff blood pressure (BP) measurement and impedance cardiography.

Situational Control and Specificity

Consistent with the situationalist perspective of cognitive behavior therapy and assessment (cf. Mischel, 1968), ATSS audiocassettes present specific, concrete stressors designed to elicit cognitions of interest relevant to a particular situation. This is similar to Wolpe's (1958) systematic desensitization, in which a list of very specific situations is constructed as a representation of the person's psychological problem (Goldfried & Davison, 1994, pp. 114–120). Thus, the ATSS paradigm provides for a high degree of situational specificity and experimental control in the assessment of cognitions. The experimenter can associate particular cognitions with specific and complex situations, make comparisons of cognitive categories across individuals or groups, and evoke cognitions in situations that are of theoretical concern but perhaps low frequency, thus complementing the random sampling of in vivo cognitions as exemplified in the work of Hurlburt (1979).

Flexibility of the Paradigm

Assessment of cognitions is limited only by the experimenter's creativity in designing provocative stimulus material and appropriate categories for analyzing content and by the ability to involve participants in ethically acceptable role playing. Such use of simulated situations, of course, has a time-honored place in behavior therapy, going back at least to Salter's (1949) proposals on pairing stressful imagery with relaxed states and to Wolpe's (1958) innovations with systematic desensitization. Situations that might be impractical or unethical to study in vivo can be simulated with ATSS, thereby permitting a study of cognitions in a much wider range of circumstances, some of which would be impossible to recreate, for example, assessment

...
of thoughts in reaction to critical comments by a deceased individual like a parent. Reliance on simulations—although it inevitably raises questions about external validity — also enables clinical researchers as well as practitioners to explore people’s reactions to abstract situations, such as being criticized or evaluated, going beyond the study and treatment of concrete and often simple phobias, like fear of nonpoisonous snakes or fear of heights. Also, nonstressful ATSS tapes allow for direct comparisons between provocative and control situations—a feature lacking in most other cognitive assessment methods.

**Empirical Findings**

In this section of the article, we review experiments that have used the ATSS paradigm, including many conducted independently of our research group (Bates, Campbell, & Burgess, 1990; Eckhardt, Barbour, & Davison, 1996; Eckhardt, Barbour, Wilson, & Davison, 1996; Eckhardt, Barbour, Wilson, Davison, Kandies et al., 1996; Laumakis, Margolin, & John, 1996; O’Brien, Balto, Erber, & Gee, 1995; O’Brien, Margolin, John, & Krueger, 1991; Warren & Smith, 1997).

**Initial Validation Studies**

The experiments described in this section provide evidence for several forms of validity for the ATSS procedure and coding systems used.

The initial experiment with ATSS (Davison et al., 1983) demonstrated its ability to elicit thoughts that could be reliably coded (a major concern with qualitative data like ours), were of theoretical interest, and could discriminate stressful versus neutral conditions. Although seldom mentioned in scientific articles, we take the liberty to say that, prior to doing this first study, we had no idea if participants would accept the experimental situation, whether they would talk much at all about what was going through their mind as our stimulus scenarios unfolded, whether the data could be coded reliably, and whether the resulting thought categories, when quantified, would show worthwhile patterns and relationships. As an example of our findings, when participants listened to the social criticism tape, they articulated more thoughts critical of the person speaking disparagingly of them, more desire to harm the speaker, more defense of the self, more resignation, and more “should” responses than they did when thinking aloud to the unknown professor control tape. These findings support the construct validity of the method. With respect to the reliability of coding, we achieved in this initial and subsequent experiments Pearson product–moment correlations or kappas ranging from .75 to .86.

In an effort to study Ellis’s (1962) cognitive theory of emotion, in the next experiment Davison, Feldman, and Osborn (1984) used a less molecular coding system. They found a greater number of irrational beliefs in the stressful than neutral situations. Participants also rated stressful situations as more anxiety provoking than neutral situations; this finding suggested good face validity for ATSS. In addition, participants high on fear of negative evaluation (FNE; Watson & Friend, 1969) and a questionnaire measure of irrational beliefs (Jones, 1968) rated the stressful ATSS situations as more anxiety provoking than did low scorers on these trait measures, providing additional construct validation for ATSS. Self-reported anxiety correlated significantly with irrational thinking in both ATSS stressful and neutral situations, providing evidence of concurrent validity.

A group effect would, we then thought, more likely be found in a comparison of a clinical sample with a control group than in a comparison of unselected students scoring high or low on FNE. Davison and Zigelboim (1987) provided this comparison plus additional construct validation for ATSS. Articulated thoughts of socially anxious participants (referred from the campus counseling center) were found to be more irrational than the thoughts of controls. As in previous ATSS studies, both groups demonstrated more irrational beliefs in their articulated thoughts in the stressful than neutral situations. Consistent findings were reported by Bates et al. (1990).

A concern with any think-aloud procedure is that participants will not or cannot report everything that is on their minds. (Endorsement methods, of course, do not permit participants to report all that they might.) Kashima and Davison (1990) therefore investigated participants’ articulations when asked to think aloud twice to the same tape. Specifically, they looked at how similar each participant’s two sets of articulated thoughts were, depending on whether they were asked to say different things the second time or just to repeat the procedure. It was found that instructions to say something different had less effect when articulated thoughts were coded functionally (viz., irrational thinking) than when they were coded in terms of surface similarity of statements. This study, then, provided evidence that the fact that participants do not verbalize everything on their minds—something that has always been self-evident—probably has little effect on functional (or, as some cognitive psychologists would say, schematic) analyses, the kind of content analyses that have been, and doubtless will continue to be, of interest in theory-driven ATSS research.

Coffman and Davison (1997) are completing a study to determine the test–retest properties of ATSS when it is used to assess cognitive and affective components of anxiety and of anger–hostility–aggression (AHA, coding used in our hypertension studies, to be described below). Another recent experiment in our lab added to the construct validity of the ATSS paradigm by demonstrating that response latencies were briefer during an emotionally provocative scenario than during a neutral one (Ring & Davison, 1996). This finding relates to physiological research using the startle eye-blink measure, whereby shorter latencies are observed under emotional arousal than during relaxation (Cook, Hawk, Davis, & Stevenson, 1991).

In addition to the experiments above, many of those described below add to the evidence for the validity of ATSS. Because these experiments related directly to specific clinical issues, we organize them topically.

**Cognitive Factors in Anxiety and Depression**

In addition to the studies described above, several others have been conducted that have tested cognitive theories and explored cognitive factors in anxiety and depression. In one study, the important question was posed: Do ATSS data relate meaningfully to overt behavior? The experiment compared ATSS assessment of self-efficacy with a behavioral assessment in speech- anxious participants (Davison, Haaga, Rosenbaum, Dolezal, &
Indices of anxiety coded from behavioral samples of speech giving in an academic class correlated with self-efficacy coded in articulated thoughts as sampled in a separate lab situation weeks later. Specifically, in stressful (but not supportive) situations, states of mind ratios (Schwartz & Garmoni, 1989) of self-efficacy correlated inversely with behaviorally rated speech anxiety and with questionnaire-based trait speech anxiety: The more anxious the participants were when giving an actual speech, the less their ATSS-measured self-efficacy was during a simulated stressful speech. This multimodal assessment of speech anxiety provided additional construct validation, relating ATSS for the first time to actual behavior outside the laboratory.

In the first ATSS study with participants having serious psychopathology, White, Davison, Haaga, and White (1992) found that in response to a simulated situation that depicted a negative occurrence (viz., that it rained on the day of the participant’s barbecue), patients diagnosed with major depressive disorder exhibited greater Beckian (Beck, 1967) cognitive bias than did patients who had other forms of psychopathology but not depression. This finding furnished both additional construct validation for ATSS and confirmation of an aspect of Beck’s cognitive theory of depression.

**Anger—Hostility—Aggression and Interpersonal Conflict**

There have been several ATSS studies of AHA and interpersonal conflict. Some of these have related AHA cognitions to physiology, whereas others have examined various cognitive aspects of marital and family conflict.

**Physiology and AHA.** Cognitive factors in borderline hypertension were explored within the context of a large clinical trial on borderline hypertension and its amelioration. Type A individuals (Jenkins, Rosenman, & Friedman, 1967; Rosenman, 1978) were found to articulate fewer self-supportive statements in a social-evaluative situation than were Type B individuals (Williams, Davison, Nezami, & DeQuattro, 1992). Findings also demonstrated that ATSS correlated with analogous paper-and-pencil measures (although we have never argued that such correlations are necessary to establish the validity and utility of the paradigm—more on this in the last section of this article). For example, ATSS measures of AHA cognitions correlated positively with scores on the trait portion of the Spielberger State–Trait Anger Scale (STAS-T; Spielberger, Jacobs, Russell, & Crane, 1983). Also, ATSS measures of self-deprecating and self-supporting thoughts correlated with FNE positively and negatively, respectively.

The principal part of this clinical trial entailed a comparison of the standard nonpharmacological medical intervention for hypertension (i.e., dietary and exercise regimens) with a treatment that also included intensive relaxation training. Consistent with previous behavioral medicine research, heart rate (HR) and BP were reduced more in patients receiving relaxation therapy than in patients who did not. An important finding for our purposes is that patients in the relaxation group had greater reductions in AHA cognitions, as measured by ATSS, and these changes correlated positively with reductions in HR and BP (Davison, Williams, Nezami, Bice, & DeQuattro, 1991). Standard paper-and-pencil measures of anger and hostility (STAST and the Cook–Medley Hostility Scale) did not detect these changes. The greater sensitivity of ATSS to cognitive changes from relaxation therapy permitted the first controlled demonstration of a link between reduction in angry thinking and diminution in BP, a correlation widely assumed to exist in the behavioral medicine literature.

In an inquiry into the possible effects of mood manipulation on AHA assessment with ATSS (Vogel & Davison, 1997), undergraduate students who were frustrated by the experimenter prior to listening to the anger-provoking garage tape did not articulate AHA thoughts any more frequently than did those students who had not been harassed, indicating that ATSS assessment of AHA is robust against experiment-induced frustration. (That the frustrated students were made more angry than the controls was evidenced by greater elevations in BP and higher scores on the Profile of Mood States anger scale.) It also turned out that in the frustrated group, the level of AHA cognitions was inversely related to baseline BP levels; that is, the lower the levels of angry thoughts expressed in ATSS—an index of anger-in (participants not expressing anger)—the higher the levels of BP. Paper-and-pencil measures of state and trait anger, however, did not pick up this relationship. The link between AHA articulations and BP is consistent with anger-in hypotheses of hypertension and provides additional evidence of construct validity for ATSS assessment of AHA.

**Marital and family conflict.** Several studies have used ATSS to study marital and family conflict and its effects on children. Warren and Smith (1997) found that men experienced as much marital distress as did women when this was coded from articulated thoughts. This finding has not previously been found with paper-and-pencil measures and indicates that ATSS may be particularly valuable in assessing sensitive issues; in other words, ATSS may be superior to endorsement methods in its ability to overcome self-presentation biases in self-disclosure. Several aspects of attribution theory were also tested, and it was revealed that individuals make attributions without being prompted to do so and that they make more negative attributions in simulated situations where behavior is depicted as blameworthy as compared with situations in which behavior is blameless.

Research by Eckhardt and associates is looking at anger, aggressive verbalization, and cognitive processing in maritally violent (MV) men in comparison with nonviolent men in a discordant marriage and nonviolent men in a satisfying marriage (Eckhardt, Barbour, & Davison, 1996; Eckhardt, Barbour, Wilson, & Davison, 1996; Eckhardt, Barbour, Wilson, Davison, Kandies, et al., 1996). Among other things, it was found that MV men articulated significantly more irrational thoughts and cognitive biases than did the other two groups of men during anger-arousing ATSS. Furthermore, MV men articulated more hostile attributional biases than did the others, and nonviolent men in a satisfying marriage articulated more anger control strategies during ATSS anger arousal than the other participants. In general, ATSS measures were better at discriminating among the three groups than were endorsement measures. The results promise to add to emerging social information-processing models of marital violence as well as to contribute to cognitive–behavioral treatment strategies for male batters.

In the first ATSS study with children, 8- to 11-year-old boys and their mothers articulated their thoughts in simulated situations that depicted parental conflict (O’Brien et al., 1991). There were several findings that shed light on the cognitive
aspects of being in a high- or low-conflict home. For example, boys from low-conflict homes were more optimistic than were boys living in physically or verbally aggressive homes. Another study involving children (ages 9–13) investigated their emotional, cognitive, and coping responses to simulated marital conflict (Laumakis et al., 1996). Among other things, it was found that boys from high-conflict homes were particularly likely to express negative evaluations and to propose intervening, thereby suggesting that boys with prior exposure to conflict may be more likely to become participants in, rather than just witnesses to, marital conflict. These and other findings highlight the usefulness of ATSS in assessing situational variables important in the impact that witnessing marital conflict has on children. The results also support the sensitization theory of the effect of exposure to marital conflict on children; that is, exposure increases their reactivity to subsequent conflict in contrast with habituation theory, which posits that they might instead get used to it and become less reactive to conflict.

ATSS was used with college students to study exposure to physical aggression in parents in terms of arousal response and cognitive processing of physically aggressive cues (O’Brien et al., 1995). Students from physically aggressive homes had higher self-reported physiological arousal and negative affect in response to simulated marital and family conflict scenarios, and they were less likely to provide spontaneous suggestions regarding how marital conflict could proceed more constructively.

Childhood aggression. The cognitive aspects of Dodge’s (Dodge & Coie, 1987) theory of childhood aggression in elementary-age children were examined in a dissertation in our lab (Bice, 1993). ATSS detected predicted differences between aggressive and nonaggressive boys in information processing, for example, more hostile attribution biases in the aggressive youngsters in a scenario in which the intentions of the provocateur were ambiguous but not when the aggressive intentions were obvious.

Smoking Cessation

Cognitive factors have been shown to be relevant to people’s efforts at smoking cessation and maintaining abstinence. In an effort to predict smoking relapse, Haaga (1989) constructed situations to simulate “smoke triggers” (e.g., feeling full after a meal). Among the many findings, ex-smokers who relapsed after 3 months had earlier articulated more thoughts of non-prompted smoking than did smokers who did not relapse. It was found also that articulated thoughts reflecting positive states of mind predicted longer periods of abstinence than did less positive states of mind (Haaga, Davison, McDermut, Hillis, & Twomey, 1993). Haaga and Stewart (1992) went on to study ex-smokers’ self-efficacy for recovery (SER) from a lapse after smoking cessation and found that ATSS measures of SER were related to maintenance of abstinence: Ex-smokers with moderate SER maintained abstinence longer than those with low SER. Such findings add to the predictive and construct validity of ATSS for studying self-efficacy as well as other cognitive factors in behavior change.

Another interesting finding from the above line of research was that several of the ex-smokers mistook the gender of audiotaped actors whose roles violated sex-role stereotypes, yet those same actors were never misidentified when portraying more traditional roles (Haaga, 1990). These gender schematic parapraxes were consistent with gender schema theory, and this finding shows that ATSS can be used for testing social psychological constructs, such as stereotyping. That these parapraxes were discovered at all demonstrates the benefits of using an open-ended response format and realistic simulated situations, such as ATSS, because these findings were unanticipated.

Psychotherapy Process

A further innovative use of ATSS involves the first direct assessment of the cognitive aspects of psychological reactance (Gann & Davison, 1997). Undergraduate students heard audiotaped simulations of both behavioral (“You will learn to relax”) and paradoxical (“Your task is to go out and make yourself more anxious”) interventions for anxiety. ATSS reactance scores were assessed by the students’ defiant articulated thoughts. These reactance scores were significantly higher for the paradoxical condition compared with the behavioral condition. No relationship was found between pretest paper-and-pencil measures of reactance and ATSS reactance scores. Post hoc findings suggest a significant gender–treatment interaction: Women scored higher than men on reactance in response to the paradoxical situation but not to the behavioral situation. This study suggests that ATSS is useful in studying psychotherapy clients’ cognitive reactions to different kinds of therapy, including the complexities and counterintuitive nature of paradoxical approaches.

Conclusions, Limitations, Possibilities

The ATSS paradigm for think-aloud assessment has been used to study cognitive factors in a number of clinically relevant contexts, including depression, hypertension, speech and social anxiety, smoking cessation, marital and family conflict, childhood aggression, and psychotherapy processes. ATSS is suitable for children, younger and older adults, patients with depression, and populations with limited reading and writing skills. It may be especially useful when relatively little is known about the cognitive terrain of interest, but it is also useful in testing previously established cognitive theories. ATSS was found in one study (Davison, Haaga, et al., 1991) to be superior to standard paper-and-pencil questionnaires in detecting cognitive changes from therapeutic intervention, suggesting its value in process research. There are also reasons to believe that ATSS is of particular benefit in studying sensitive topics and in reducing or avoiding presentation biases in self-disclosure (e.g., Warren & Smith, 1997). A variety of ATSS coding strategies have been reliably established, and the studies described above furnish evidence for its face, concurrent, predictive, and construct validity. In general terms, think-aloud procedures like ATSS allow researchers to assess issues fundamental to cognitive behavior therapy, other cognitively oriented therapeutic approaches, and cognitive theories of psychopathology. Assessing which aspects of environmental and internal stimuli an individual attends to and the meaning derived from attention to this subset of stimuli can add to our understanding of how a particular individual organizes and perceives his or her environment.

It should be made explicit that ATSS, at least as applied so far, cannot justify cause–effect statements regarding cognition,
on the one hand, and emotions and behavior, on the other. The most one can determine, as seen in the studies reviewed in this article, is whether there are relationships between what people articulate about their thoughts and the situations they find themselves in as well as other aspects of their behavior and personality. The determination of such complex relationships is, of course, a necessary condition for addressing cause-effect issues because while correlation does not imply causation, the reverse certainly holds. We believe that the ATSS paradigm offers a useful way to provide data that bear on whether people's cognitions affect the way they feel and behave in a wide variety of situations—many of them too abstract and complex to examine in vivo. ATSS can not only provide information on cognitive correlates and causes of psychopathology but also furnish data relevant to the choice of therapeutic procedure. With respect to the latter, a cognitive intervention might be more effective if it is directed to certain undesirable or faulty thought processes than if it is applied to patients whose distorted thinking does not match the particular intervention. For example, teaching patients with depression not to overgeneralize in their information processing may work better if, in fact, their depression is associated with overgeneralization. (Of course, the substitution of an adaptive mode of thinking might be effective with a wide range of existing cognitive processes, whether they match the intervention or not.)

In studying psychotherapy process, we can use ATSS to help address a number of questions. For example, it is conceivable that an intervention like rational-emotive behavior therapy changes not what people tell themselves—Ellis's (1962) assumption—but how they feel about the very same self-statements. A cognitive therapy, then, might leave intact cognitions that are associated with negative affect but might bring about improvement through an "unhooking" of the undesirable emotion (or behavior) from the negative self-talk. The effects of psychoactive drugs could also be studied in this way; to wit, does a tranquilizer calm down an anxious person by changing the way he or she construes the world or does it make him or her less upset about the way things are construed? Furthermore, is postdrg maintenance of therapeutic change influenced by one or the other of these two putative modes of action? Perhaps those drugs from which patients can be weaned without losing therapeutic benefit differ from drugs that must continue to be taken because the former allows or encourages a change in self-talk, something that may "belong to" a person more than does a drug-produced separation of cognition from affect (cf. Davison & Valins, 1969, on self-attribute vs. drug attribution in the maintenance of behavior change). A fine-grain analysis of thought through ATSS might address such issues of possible importance for understanding the processes by which different interventions effect therapeutic change.

The issue of convergent validity of diverse methods is longstanding in psychological assessment, and the problems relating endorsement methods with production approaches like ATSS have been recently reviewed and analyzed by Chamberlain and Haaga (in press). As the present article indicated, ATSS sometimes correlates with other, more standard assessment methods like questionnaires, but sometimes it does not. As noted earlier, we have never taken the position that the validity or utility of ATSS relies on such convergence. We hope we have shown how different ATSS is in its approach to cognitive assessment. In this instance, method variance is not undesirable "noise" to be eliminated. Rather, differences in findings that "ought to" be consistent because the concepts being studied are the same or similar can and perhaps should, we would argue, be viewed as a positive: Different methods can contribute different information about the same concept. Multitrait-multimethod approaches to the establishment of validity have wisely been advocated for many years in the clinical assessment literature (e.g., Campbell & Fiske, 1959). Certainly ATSS data should be related to other methods of cognitive assessment, particularly to other production tools such as the Dysfunctional Thoughts Record (Beck, Rush, Shaw, & Emery, 1979), endorsement methods such as the Automatic Thoughts Questionnaire (Hollon & Kendall, 1980), the in vivo thought sampling of Hurlburt (1979), and the thought-listing technique of Cacioppo et al. (1979, see also Cacioppo, von Hippel, & Ernst, 1997). As efforts are made to establish convergent validity, we would suggest that the unique perspectives that certain assessment approaches offer not go unappreciated.

It cannot be emphasized enough that ATSS is not an instrument or even a specific method. (Davison has had to point this out more than once to editors who have put together compendia of assessment methods.) The use of the word paradigm is intentional. There is no set way to present scenarios (audio, video), no set number of segments per scenario, and no standard ways to content analyze the data. Indeed, we are not aware of direct comparisons, within the same data set, of different coding schemes that purport to get at the same constructs. For example, would the molecular coding system used in the first ATSS experiment (Davison et al., 1983) relate to the more molar system used in the second study (Davison et al., 1984)? A few studies have used the same content analytic system (e.g., Davison, Williams, et al., 1991; Vogel & Davison, 1997; Williams et al., 1992), the Davison, Williams, et al. report using posttreatment data as well as the pretreatment data that the Williams et al. article focused on, but this overlap is more the exception than the rule. An advantage of ATSS, however, is that the raw data are amenable to multiple content analyses, so that archived data are available for such comparisons (Haaga & Allison, 1994).

An additional consideration on coding was recently elaborated by Chamberlain and Haaga (in press). They rightly pointed out that data from production approaches like ATSS, coded by raters other than the participants who provided the think-aloud data, may have meaning ascribed to them that participants did not intend. There is an interesting irony here. As argued elsewhere (Davison & Neale, 1998, p. 537), cognitive therapy is inherently phenomenological; that is, it takes as a given that an individual's particular construction of a situation is an important determinant of his or her reaction to it. Yet, most production assessment methods have independent coders give meaning to participants' think-aloud verbalizations by applying experimenter-constructed content analytic schemes to the data. Such meanings may well be different from the participants'. Yet, it is uncertain whether having participants code their own think-aloud data, as suggested by Cacioppo and Petty (1981) for thought-listing productions, is necessarily the best strategy. For one thing, the categories one might apply to think-aloud data might require extensive training that would not be feasible or possible to provide to experimental participants. In our experience, coding for such concepts as irrational thinking
or Beckian (Beck, 1967) cognitive biases like selective abstraction or maximization–minimization requires many hours of intensive instruction and supervision before coder understanding and intercoder reliability are achieved. We would agree with Chamberlain and Haaga (in press) that the jury is still out on this important and intriguing problem.

Thus, the assessment characteristics and limitations of ATSS are far from resolved. More research is needed regarding such issues as what its psychometric properties are, whether simulated situations evoke the same cognitive processes as real-life events, and whether and how ATSS relates to other cognitive assessment techniques. Nonetheless, the paradigm is versatile and adaptable, limited only by ethical constraints and by the creativity of the experimenter interested in what is on people’s minds as they negotiate their way through life.

References


Received January 3, 1997
Revision received February 24, 1997
Accepted March 17, 1997