

Learning How To Rate Video-Recorded Therapy Sessions: A Practical Guide for Trainees and Advanced Clinicians

Leigh McCullough

Modum Bad Research Institute, Vikersund, Norway and
Psychotherapy Research Program at Harvard Medical School's
Beth Israel-Deaconess Medical Center, Boston, MA

Maneet Bhatia

McGill University

Pal Ulvenes and Lene Berggraf

University of Oslo

Kristin Osborn

Psychotherapy Research Program at Harvard Medical School's
Beth, Israel-Deaconess Medical Center, Boston, MA

Watching and rating psychotherapy sessions is an important yet often overlooked component of psychotherapy training. This article provides a simple and straightforward guide for using one Website (www.ATOSTrainer.com) that provides an automated training protocol for rating of psychotherapy sessions. By the end of the article, readers will be able to have the knowledge to go to the Website and begin using this training method as soon as they have a recorded session to view. This article presents, (a) an overview of the Achievement of Therapeutic Objectives Scale (ATOS; McCullough et al., 2003a), a research tool used to rate psychotherapy sessions; (b) a description of APA training tapes, available for purchase from APA Books, that have been rated and scored by ATOS trained clinicians and posted on the Website; (c) step-by-step procedures on how ratings can be done; (d) an introduction to www.ATOSTrainer.com where ratings can be entered and compared with expert ratings; and (e) first-hand personal experiences of the authors using this training method and the benefits it affords both trainees and experienced therapists. This psychotherapy training Website has the potential to be a key resource tool for graduate students, researchers, and clinicians. Our long-range goal is to promote the growth of our understanding of psychotherapy and to improve the quality of psychotherapy provided for patients.

Keywords: ATOS scale, www.ATOSTrainer.com, psychotherapy training

Receiving feedback on performance is becoming increasingly recognized in many fields. How many times have we read in the sports section of the newspaper a player or coach after a tough game saying, “we had a rough night but we’ll go back and watch the film to see where we went wrong,” or when planning for the next opponent, “we’re going to watch film on their players to better understand how they implement their strategies.” Watching video plays a crucial role in the development of athlete’s skills,

assisting both player and coach with learning team strategies, improving team chemistry and identifying specific areas in which the player needs to further improve and develop. The key is that continuous practice and observation improves performance enabling you to become a better athlete. With repetition and proper technique, a basketball player perfects the art of shooting; a baseball player improves his hitting stance; a golfer learns the proper technique and form to drive the golf ball properly. These skills are enhanced by the athlete and coach sitting together and analyzing videotapes to decipher ways the athlete performs each task, and how they can improve. Performance video-recordings allows for an objective evaluation of techniques and progress and a platform for the coach to provide individualized feedback for the athlete.

Psychotherapy, like art or music, is taught in the master-apprentice model, but historically the apprentice never could see the master practice! Now with the use of recorded sessions, the psychotherapy trainee can see the “masters” at work, and can gauge their own progress in skills based on that. As such, there are now parallels between sports training and learning psychotherapy. Therapists, like athletes, have supervisors as “coaches,” and being able to videotape and rate therapeutic sessions is a valuable component of supervision and training (McCullough, 2003). Viewing recorded sessions is an objective method to evaluate therapeutic progress and to explore specific areas where the therapist struggles. The therapist and supervisor can analyze specific applied

Leigh McCullough, Modum Bad Research Institute, Vikersund, Norway and Psychotherapy Research Program at Harvard Medical School's Beth Israel-Deaconess Medical Center, Boston, MA; Maneet Bhatia, McGill Psychotherapy Process Research Group (MPPRG), McGill University, Montreal, Quebec, Canada; Pal Ulvenes and Lene Berggraf, University of Oslo, Oslo, Norway; and Kristin Osborn, Psychotherapy Research Program at Harvard Medical School's Beth, Israel-Deaconess Medical Center.

Special thanks to Jeremy Osborn for developing and building the www.ATOSTrainer.com Website in such short notice. This project would not have been possible without his expertise and hard work. Special thanks to Jonathan Petraglia (McGill University) and Shafik Sunderani (McMaster University) for their comprehensive and extremely valuable comments of earlier versions of the manuscript.

Correspondence concerning this article should be addressed to Maneet Bhatia, MA, McGill Psychotherapy Process Research Group (MPPRG), McGill University, 3700 MacTavish Street, Montreal, Quebec, Canada H3A 1Y2. E-mail: maneet.bhatia@mail.mcgill.ca

techniques, the manner in which they are delivered, and the dynamic relationship that exists between therapist and patient. The objective is that continuous observations with a master therapist, and practicing new skills, help the therapist master a new behavior.

It is our belief and experience that by having therapists and students alike engage in the process of videotaping and rating their therapeutic sessions, and comparing those ratings to expert ratings, that they come to a deeper appreciation and understanding of the psychotherapeutic process and ultimately become better clinicians. Therefore, the aim of this article is to teach the reader how to effectively rate the quality of psychotherapy sessions on important common factors in psychotherapy. This article will provide you with the “nuts and bolts” of rating your tapes, and demonstrate this process by applying it to a scale created for the *sole* purpose of coding psychotherapy sessions.

The APA “Master Therapist” Series of Recorded Sessions

In the development of the online rating system, we have been hampered for many years by not having recorded sessions that could be publicly shown. APA Videos has developed such a series of recorded sessions, with many different therapists, therapies, and many different types of patients. The DVDs can be obtained from APA Video Series (<http://www.apa.org/pubs/videos/index.aspx>). The therapy sessions deal with real people and real problems. The individuals who have volunteered to be recorded have agreed to allow their session, or sessions, to be sold and used for research

and training in psychotherapy. In return they have a free session with a “master therapist” and are given a recording of the session.

We purchased these APA videos to train our research assistants for the Process Mapping Program at the Modum Bad Research Institute. (This is the largest psychotherapy video analysis ever undertaken. Over 1,500 tapes will be analyzed with the Achievement of Therapeutic Objectives Scale [ATOS] scale.) As a result, we have generated “Expert Ratings” so we can establish reliability when student trainees start generating ratings. During this process, we realized that we had created a training method we could share with others also interested in examining psychotherapy. Currently, there are ATOS ratings available for nine APA sessions available on the ATOSTrainer Website for you to score and compare with expert ratings; (1) Leigh McCullough, (2) Nancy McWilliams, (3) Paul Wachtel, (4) Jeremy Safran, (5) Jon Carlson, (6) Scott Miller, (7) Reid Wilson, (8) Gordon Wheeler, and (9) Marvin Goldfried. (See Table 1 for therapy model and order information). These sessions represent many different types of theoretical orientations, including psychodynamic, (long and short term), cognitive, interpersonal, Adlerian, and behavioral. All were rated with the ATOS scale by expert raters so that a person who desires training in evaluation of sessions can compare their ratings to the expert criterion. There are also many different types of patients, including personality disorders, obsessive–compulsive disorders, anxiety disorders, and depression.

These nine initial tapes are just the tip of the iceberg. We also hope that more instruments, such as therapeutic alliance measures, and the Psychotherapy Q Sort will be added. Now, because of the

Table 1
APA Systems of Psychotherapy DVDs Used for ATOS Training

Therapist	Type of therapy and title	Item # and ISBN #	Publication date and minute	Cost
1. Leigh McCullough	Affect focused short term psychotherapy: “The Worried Mom”	Item #: 4310728; ISBN: 978-1-59147-301-5	June 2005; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
2. Nancy McWilliams	Psychoanalytic therapy	Item #: 4310811; ISBN: 978-1-4338-0123-5	September 2007; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
3. Paul Wachtel	Integrative relational psychotherapy	Item #: 4310793; ISBN: 978-1-59147-801-0	March 2007; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
4. Jeremy Safran	Relational psychotherapy	Item #: 4310846; ISBN: 978-1-4338-0328-4	May 2008; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
5. Jon Carlson	Adlerian therapy	Item #: 4310721; ISBN: 978-1-59147-265-0	February 2005; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
6. Scott Miller	Client directed outcome focused psychotherapy	Item #: 4310729; ISBN: 978-1-59147-303-9	June 2005; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
7. Reid Wilson	Obsessive compulsive disorder	Item #: 4310711; ISBN: 978-1-59147-217-9	October 2004; 100+ min	List price: \$99.95; member/affiliate price: \$69.95
8. Gordon Wheeler	Gestalt therapy	Item #: 4310703; ISBN: 978-1-59147-177-6	June 2004; 100+ min	List price: \$99.95; Member/affiliate price: \$69.95
9. Marvin R. Goldfried	Cognitive-affective behavior therapy	Item #: 4310802; ISBN: 978-1-59147-812-6	June 2007; 100+ min	List price: \$99.95; member/affiliate price: \$69.95

extraordinary contribution by APA Video series, we were able to proceed with the automated, online rating of sessions using ATOS.

The ATOS Scale of Common Factors in Psychotherapy

The ATOS scale is a process instrument originally created to assess patients' degree of absorption, assimilation, or achievement of specific treatment objectives characterized as essential change mechanisms in short-term dynamic psychotherapy (STDP; McCullough et al., 2003a; Valen, in press). STDP theory states that psychopathology is rooted in psychodynamic conflicts or conflicts about feelings, where defenses and anxieties block the experience and expression of adaptive affects. Using learning theory language, these conflicts can be described as "fear of feelings" or Affect Phobias." Usually, a conflict or phobia is centered on one core affect, such as grief, anger, or self-compassion in their adaptive and activating modality. The goal of successful psychodynamic psychotherapy then is to help patients' decrease the inhibition (e.g., anxiety, guilt, shame) of those "adaptive activating feelings." The decrease in anxieties thus increases the tolerance of the experience of the affect. It also makes the patient better able to express the previously warded off affects with appropriate modulation. As such, better tolerance and expression is achieved through a process of (a) recognizing and relinquishing maladaptive or defensive behaviors, (b) desensitization of warded-off or blocked affects through exposure to conflicted feeling, and (c) alteration of maladaptive conceptions of self and others (McCullough et al., 2003b).

Although the ATOS scale was originally designed to assess factors central to change in dynamic forms of psychotherapy, the resulting seven objectives were operationally defined in behavioral language and overlapped significantly with well-established "common factors" empirically shown to play a role in therapeutic change (McCullough et al., 2002; McCullough et al., 2003b). For example, being able to recognize defensive or maladaptive behaviors can be seen as a form of the common factor "Insight" (See Kallestad, Valen, Svartberg, McCullough, Stiles, in press); the desire to give up maladaptive or defensive behavior can be seen as a form of the common factor "Motivation." The degree of affective arousal and experiencing is akin to the common factor "exposure to (adaptive, activating) feelings," and the capacity to express feelings in appropriate ways could be seen as a form of the common factor "new learning." "sense of self" can be seen as the common factor "improvements in self-image" and "sense of others" could be conceptualized as "improvements in relationships."

The ATOS scale consists of identifying the predominant affect in the session, and rating the seven subscales, each to be rated on a 1–100 scale. (For examples of how, and a more detailed demonstration of how the ATOS scale can be adapted to different therapy modalities, see Table 2). As demonstrated above with the APA sessions, the ATOS scale has been used to analyze a wide range of psychotherapies. A recent study examined the ATOS inter-rater reliability and the sensitivity to change in STDP and cognitive therapy (CT). The results found good-to-excellent inter-rater reliability and in both STDP and CT, there was adequate sensitivity to change in theoretically consistent ways in both therapies (Valen, in press). These findings provide additional evidence for the psychometric de-

pendability of the ATOS scale and its utility when comparing dynamic and cognitive therapies. We believe that the ATOS scale has the potential to evaluate a broad range of other therapies, as well.

The "Nuts and Bolts" of Watching and Rating Recorded Therapy Sessions Introducing www.ATOStrainer.com

The Website www.atostrainer.com is designed to facilitate Internet-based watching and rating therapy sessions, using the ATOS scale and is accessible to all for free. The current site has expert rated ATOS scores for nine APA DVDs. Our hopes are to add additional expert rated ATOS scores for other DVDs in the future. It is important to reiterate that the Website does not contain the actual APA videos and they must be purchased from the APA Website given earlier in this article. Additionally, we also are looking forward to the possibility that others scales will be added.

Steps for Rating

We now turn our attention to the "nuts and bolts" of watching and rating recorded sessions. In our experience the richest learning occurs when raters watch tapes together and are able to discuss ratings and their rationales for them. With more raters involved it allows for the opportunity to integrate each individual's perspective on the therapist-patient interaction and helps each raters' see the therapy unfold in a way that they may not have otherwise seen. Raters' may not agree with one another's interpretations, but having disagreements and passionate debates among one another can provide for valuable learning. The steps that follow apply whether you are rating tapes individually or in a group setting and this can depend on the resources available to you. The steps are broken down into phase 1 and phase 2. Both phases must be followed to fully engage in the rating process.

Phase 1: Watching Sessions and Using the ATOS

Step 1: Purchase APA sessions or record your own. Purchase one of the APA DVDs that were discussed earlier in the article. These DVDs can be purchased from the APA Website at www.apa.org (Table 1). We have generated criterion "Expert Ratings" on each of the ATOS subscales for each of these DVDs. You then will be able to compare to your own ATOS ratings with the "Expert Ratings."

Another possible way to learn the ATOS is to rate recordings of your own sessions to begin practice, using the ATOS scale. You would not have "Expert Ratings" to compare them with, but it will help you become familiar with the ATOS procedures, and it will give you an idea of how you are doing in your sessions. The ATOS scale is strongly grounded in observable behaviors, and can be largely self-taught. So if you attend carefully to the patient's behaviors, and carefully read the examples at each ATOS level, (including the level both above and below your rating to make sure of your rating level) you will be able to get a fairly good sense of how you are doing even without expert ratings to guide you.

Step 2: Read the ATOS manual. Download the full ATOS manual at <http://www.atostrainer.com/#manual> by clicking on pdf link under "The ATOS Coding Manual" heading.

Table 2
Description of The ATOS Scale

Treatment objective	STDP definition	CT definition	Ratings are based on	Examples
Insight	The patients' degree of recognition and understanding of their own pattern of defensiveness, anxieties and feelings	How clearly patients can identify or understand their maladaptive cognitions, cognitive schemas or compensatory strategies in general.	(a) How clearly and fully the patient describes maladaptive patterns of thoughts, feelings and/or behaviors with explicit examples, and (b) the degree of ability to state why and how maladaptive/defensive patterns began and are maintained (secondary gain, meaning, causes, and with whom).	Clear and comprehensive descriptions of maladaptive behavior patterns, with links to origin in past as well as reasons for these maladaptive patterns.
Motivation	The patient's degree of motivation to change or relinquish destructive or defensive behavior	How prepared the patient is to give up the maladaptive schemas or compensatory behaviors.	(a) The patient's degree of motivation to give up maladaptive patterns of thoughts, feelings, and/or behaviors, and (b) the degree of dislike, undesirability or sorrow related to the cost of defenses or maladaptive behaviors.	Grief over cost of defenses, or openness to change.
Exposure (adaptive affect)	The patient's intensity and duration in experiencing activating affects such as adaptive versions of grief, anger, interest, joy and fear	Affect arousal is not a primary focus, but is often an indication of activation of negative core schemas.	(a) The intensity of arousal of adaptive affect as shown in the patients' vocal tone, facial expression, nonverbal or charged verbal statements during a session, (b) the duration of the affective arousal and (c) subsequent expressions of relief	Grief that cause improved acceptance of a loss, anger that sets appropriate limits, interest that generates enthusiasm and motivation, joy that causes peace, tranquility or relaxation (blood flow to the extremities) or fear that causes escape from harm. Examples are spirited but well-controlled and well-integrated communication of thoughts and feelings, assumed in STDP theory to bring a sense of relief or satisfaction by feeling understood and whole. That is authentic behavior and emotional expression with others, reports of adaptive crying when alone (if not to avoid doing so with others), adaptive self-care or self-talk when alone.
New learning	The patient's ability to express thoughts, feelings, wishes, and needs accurately and adaptively in real-life situations	The patient's ability to generalize and express thoughts, wishes and needs in between-session situations.	(a) Appropriate, adaptive interpersonal face-to-face expressions of thoughts and feelings, and (b) the degree of relief or satisfaction versus discomfort in the action of expression.	Examples include trembling, tension or shifting (anxiety), blushing or head down (shame or guilt) or wincing, groaning or whimpering (pain). Other examples common to one or more of the above include hesitation, looking away and guardedness.
Inhibition (inhibitory affect)	Levels of bodily expressions of anxiety, guilt, shame and pain. The reduction of inhibitory affects is a key component of each of the other six treatment objectives; i.e., treatment seeks to reduce the level of inhibitory affects (e.g., anxiety, guilt, shame, etc.) associated with defenses, affects, and relationships.	Same as STDP	(a) The degree of inhibition or overall intensity of the observable anxiety, guilt, shame or pain as shown in verbal report, vocal tone, nonverbal behaviors and physiological signs of inhibition. Examples include trembling, tension or shifting (anxiety), blushing or head down (shame or guilt) or wincing, groaning or whimpering (pain).	

Table 2 (continued)

Treatment objective	STDP definition	CT definition	Ratings are based on	Examples
Sense of self	Assesses the patient's degree of adaptive positive feelings about the self, including healthy interests, excitement, pride, self-compassion, self-care, and acceptance of own limitations or realistic negative qualities of the self, but without grandiosity or devaluation of self	Reflects self-esteem as expressed in positive self-talk.	The patient's verbal report of adaptive self-image, in terms of (a) the degree of experience of self-compassion, self care or value as a human being, (b) the degree of adaptive pride in positive qualities (self worth, self esteem, competence) and (c) the degree of ability to compassionately acknowledge and accept one's limitations or realistic negative qualities of the self.	Examples include healthy pride in own strengths (not grandiose), great sense of self-compassion and self-acceptance.
Sense of others	Assesses patient's capacity to view others in a realistic light (without devaluation or idealization)	The scale captures improvements in interpersonal relationships.	The patient's report of adaptive and realistic images of other people, in terms of (a) the degree to which the patient can acknowledge or respond to others' positive qualities and (b) the degree to which the patient can acknowledge and set limits around destructive or (realistic) negative qualities in others.	The ability to integrate positive and negative qualities of others.

Step 3: Watch a 10-min segment. Start watching the video of your choice in 10-min segments. Sometimes the time is stamped on the DVD, but if not, you can check the minutes on the DVD recorder or computer where you are playing it. While you are watching the 10-min segment, write down key comments, observations, and behaviors that are relevant to that segment. This will help you make decisions on ratings. For example, (1) any observable or verbally reported defensive behavior, (looking away, changing the subject), defensive statements (I'm not mad. I'm not an angry person.), defensive feelings (I'm so frustrated, I can't stand this anymore!); (2) any observable or verbally reported signs of anxiety, guilt, shame, or pain, (3) any adaptive affects, either verbal (I am so sad.) or nonverbally expressed (tears in eyes, hand forming into a fist) or (4) any comments referring to self or others.

Step 4: Stop the recording and rate. Once 10 min are complete, stop the DVD, and identify the adaptive affect that is exposed or needs to be exposed in the 10-min segment. In some cases there might be more than one adaptive affect. In this case, we suggest you score only the predominant affect in the segment. However, you can practice scoring the segment for each feeling. For instance, the patient may have experienced and expressed both anger and sadness so you score for anger and then go back and score for sadness. The reason we score ATOS only in relation to specific affects is because defenses and inhibitions for anger can be different than those for sadness or self-compassion.

Once the adaptive affect is identified then you are ready to score based on your notes and observations. If you are in a group setting, you can discuss your ratings with the group and come to a consensus rating. We strongly recommend that you get colleagues to join you, because it is fun and enlightening to discuss cases this way!

Phase 2: Entering Your Ratings on www.ATOSTrainer.com

Phase 2 will require the user to visit www.ATOSTrainer.com and will outline the steps the user needs to take in order to log in and enter their ratings for the segments they watched in phase 1 of this process.

Step 1: Creating your username and password. Go to www.ATOSTrainer.com. If you are a first-time user to the site you will need to create a new account and log in to the site. In order to create a new account, the user must click on the "Register" tab on the top right part of the home page. Once you click on "Register" you will be asked to provide your e-mail address, password, and agree to the "Terms of Service" which can be read by clicking on blue "Terms of Service" link. Once you have completed these steps you will be logged into the site. User names are always individuals e-mail addresses. If a user misplaces or forgets their password they can click on the "Forgot Password" link on the top right portion of the home page. Here users will be asked enter their e-mail address and they will receive an e-mail shortly after providing them with the steps needed to create a new password. For users who have already created an account, they will need to enter their user name and password and can move onto step 2 directly.

Optional Step: Video Tutorial

The home page of www.ATOSTrainer.com features a 5–6 min video, which demonstrates the features of the application. This

video is targeted primarily toward new users who are using the site for the first time, although it would also be beneficial to any visitor who might be curious about the site.

Step 2: Choosing session to rate. Once logged in, the user clicks on the “Choose Session to rate” tab. In this section they are presented with a list of the therapy sessions available for coding. Each session corresponds to the available APA tapes for purchase outlined in Table 1 of this article.

Step 3: Choosing segment to rate. Once the user clicks on the session they wish to rate they will be presented with a list of five segments. The most common scenario would be for a user to start with segment 1 and proceed sequentially, although this is not required.

Step 4: Choosing training mode or reliability mode. Before they can proceed with the coding process, the user is required to choose whether they are rating in “training” mode or “reliability” mode. In training mode the user’s scores are never saved, whereas in “reliability” mode the user’s scores are always saved and are used to calculate a final reliability rating. The purpose of having these two modes is to provide different levels of accountability in the academic setting. For example, a teacher may ask a student to code tapes in reliability mode as a form of a “test.” (This process is explained in more detail in step 7).

Step 5: Entering scores. Whether they are in training mode or reliability mode, the user now begins to fill in their rating scores for the various objectives: Insight, Motivation, Adaptive Affect, and Inhibitory Affect. Repeat these procedures for each segment you rate. The last segment (segment 5) of each therapy session will ask you to rate, “New Learning,” “Sense of Self,” and “Sense of Other.” Since there is less fluctuation on these three categories from segment to segment these three objectives are only rated once (Ulvenes, personal communication, January 18, 2011; McCullough, 2003).

For each segment, the user will need to choose an adaptive affect for that segment (e.g., Anger, Sadness, Grief, etc.) by scrolling through the list in the “Select Adaptive Affect for this segment” tab located just above the objectives. If users forget to select an adaptive affect, they will be prompted when pushing the “Calculate” button.

Step 6: Calculating ratings. Once the user has filled in their ratings for the objectives they press the “Calculate” button. The calculate function displays the expert ratings next to the user’s ratings as well as the difference between the two. In this manner, the trainee can immediately gauge how accurate their ratings were. Additionally, to help the user better understand how the expert ratings were determined, justifications in the form of a paragraph

of explanatory text are given (See Table 3 for visual representation of user ratings compared to expert ratings and the justifications). As seen in Table 3, the justification for the expert rating is divided into two categories: *Not higher because* and *Not lower because*. These help the user get a better sense of the rationale behind the expert rating.

Step 7: Reliability mode. As noted in step 4, the user must choose between “training” mode and “reliability” mode. When the user is in reliability mode, their ratings for each segment begin to be saved into the application’s database. “Reliability Mode” is designed to test how well the user’s ratings match the expert ratings and if they can be considered a “reliable” rater.

In an effort to determine how reliably a user would rate future segments, ATOS trainer uses ICC (intra class correlation) to determine a user’s reliability. ICC is a form of statistical analysis, which provides the chances a user would agree with the expert ratings. ICC is measured from 0 to 1, with higher numbers representing a higher chance of reliability. The scale used is based on recommendations of Fleiss and colleagues (Fleiss, 1981; Fleiss & Cohen, 1973; Shrout & Fleiss, 1979) and provide referents to the magnitude of standard estimates of reliability, Kappa or ICC, in the following ranges: <.40 = poor; .40–.59 = fair; .60–.74 good; >.74 excellent.

Users are given ratings for each objective on a given segment. ICC is an inference-based estimation, and for the purposes of ATOSTrainer.com, a .6 rating or above is considered to be a successful average. For statistical reliability purposes, the user does not receive an ICC rating until they have rated a minimum of 10 segments. Within the application, new users who have not yet rated 10 segments receive a small alert whenever entering “Reliability Mode.” This alert disappears after 10 segments (or more) have been rated.

Step 8: Displaying reliability scores. After the user has rated at least 10 segments a new section of the page is exposed on “Choose a Session” page. This section of the page informs the user of their current ICC ratings in two categories: As a subscale of each objective and as an overall number between zero and one.

Subscale ratings: Having the ICC rating for each subscale is helpful for the user to know which scales they are struggling with and therefore where they need to put in more effort to get reliable.

Overall rating: The overall rating allows the user to view and track their overall reliability for *all* objectives in *all* the segments rated up to that point.

Step 9: Saving and tracking progress. One feature of the ATOSTrainer worth noting is the manner in which the user’s segment ratings (and by extension their ICC scores) are cumulative

Table 3
Example of a Rating Matrix On www.ATOSTrainer.com

Objective	Your rating	Expert rating	Difference	Justification
Insight	33	43	10	Not higher because: there are no clear examples of the maladaptive behavior, no past-present links, no mention of why she has difficulty with focusing on feelings or what advantages she has of being intellectual. Not lower because: patient on her own begins to describe maladaptive patterns; “I’ve been mental always”. Patient describes it as new and hard for her to focus on feelings. However, she does not seem to link her present maladaptive patterns with her early experiences.

and saved into the user's profile. Each time a user rates a new segment in "Reliability Mode," these ratings are factored into the overall pool of ratings.

Additionally, at any point, after rating a segment, a user may click on the "Clear Scores" button and remove these scores from the database. This action will cause the ICC to be recalculated. This feature allows users (such as a training assistant learning how to code tapes) to work toward a goal of reliability without being "penalized" if their early ratings are poor. The ability to delete ratings for any given segment allows a user to narrow the scope of which of their rating scores are being used in the ICC calculation.

Video Training Strategies for Beginning Therapists

For many graduate students, there is a disconnect between research and practice where their research interests do not have clinical value or their passion for clinical work does not translate into research interest. Additionally, in the specific realm of psychotherapy research and practice, there is a strong movement toward creating psychotherapy research and practice networks with the goal of bridging the divide between both sides. Using the ATOS scale provides practitioners with an opportunity to improve their clinical work as well as providing them with opportunities in psychotherapy research that help reduce the gap between these paradigms.

The second author (Maneet Bhatia) is an advanced doctoral student in Counseling Psychology. He began learning about the ATOS scale during his Master's degree. He began rating tapes using an Internet-based site for training on the ATOS scale as well as rating "gold standard" tapes along with his colleagues at McGill University. Eventually, the process culminated in their research group (McGill Psychotherapy Process Research Group) publishing a single-case study exploring the adaptive changes the patient made on the ATOS scale and other key assessment measures during the early phases of treatment (Bhatia et al., 2009). This experience was valuable from both a research and clinical perspective.

This is an example of how watching and rating therapy sessions provided students an opportunity to learn about psychotherapy process research, become trained on a psychotherapy rating scale, conduct theses and publish articles, as well as provide them with the opportunity to watch an expert clinicians conduct psychotherapy.

The first strategy that beginning therapists should employ in order to fully benefit from the video training is to read the ATOS scale thoroughly and become familiarized with the rating system for each treatment objective. Second, the beginning therapists and raters should have a good understanding of the therapeutic orientation they are planning to rate. The Website www.ATOSTrainer.com provides expert rated ATOS scores of APA DVDs for many theoretical orientations (Table 1), and even though the ATOS scale measures treatment objectives that are common to many treatments, it is important to have a basic understanding of the principles, techniques, and therapeutic objectives of the therapy you plan to rate.

A key benefit for the beginning therapist that watching and rating videotapes of different treatments and key psychotherapy variables provides is a deeper understanding and appreciation of the art and science of psychotherapy that could not possibly be

captured through reading a case vignette or engaging in role-playing. For example, the beginning therapist observes the importance of: senior therapists demonstrating the structured use of specific therapeutic interventions at specific times (e.g., challenging a patient's defense of intellectualizing when the therapist brings up emotional topics, or the use of Socratic dialogue when exploring the evidence for or against a particular thought); the therapist being able to make an intervention based on the patient's nonverbal and body language (e.g., the therapist telling the patient, "do you notice that every time I ask you about your feelings toward your father your hands begin to clench into a fist?"); and the ebb and flow of therapy that makes psychotherapy such a dynamic process.

Another benefit of watching and rating recorded sessions in comparison to reading a case vignette or transcript is that it brings the learning experience for trainees "to life," in that you get to see how therapy *actually* occurs. When reading an article, textbook, or transcript you do not capture the nuances of therapy such as, what happens when a therapeutic intervention goes wrong, or when the therapist makes a "mistake," or when the patient responds a certain way—for example, defensive, passive. For the most part, in training manuals we are exposed to the optimal ways therapeutic interventions are delivered. There is an assumption that implementation of that intervention and the outcome of that intervention is a static process, that is, intervention X leads to outcome Y. Clinical experience is not indicative of this, and for beginning therapists watching actual therapy within this context allows for a more balanced understanding of the ups and downs and intricacies of the therapeutic process. The key training benefit occurs through the observation process, which is anchored by the ATOS scale.

One useful training technique is to have the trainees stop the tape after a patient statement prior to the expert making an intervention and provide their own intervention first, almost as if he or she were there in the room with the patient, imagining what to say or do next, and gaining valuable insight each time the therapist and patient react a certain way. Also, after the expert has made his or her intervention they can stop the tape and compare their own interventions with that of the expert and discuss. If they are in a group training session, each individual rater can share their intervention with others in the group providing rationales as to why they would have used one specific intervention over another.

Another useful training strategy is *paying attention and observing* how the patient is reacting to the therapist interventions. The patient's responsiveness to the therapist captures the essence of the ATOS scale and is an immensely important aspect of watching and rating recorded sessions. As a rater, this provides key observable markers for the trainee and helps them answer certain questions such as "How does the patient respond when the therapist gently guides the patient to face difficult feelings with the therapist?"

An important benefit of this training strategy is that it provides crucial training for raters to be able to note subtle shifts in the patients' responding when, later, they are conducting their own therapy sessions. For example, the novice therapist may recognize that when a patient is avoiding eye contact, or smiling nervously or fidgeting in-session that this is a sign of anxiety and distress related to the topic at hand or based on an inquiry they have made. In the past, the novice therapist may not have marked these subtle shifts as being therapeutically valuable. Using the ATOS scale as a guiding mechanism while watching multiple sessions and ther-

apeutic cases, the novice therapist is able to acquire the skill to decipher more readily, and clearly, patient responses. Like practice in sports, repetition of an act improves performance and increases your understanding of your strengths and limitations. The more exposure novice therapists have to different therapeutic situations the better equipped their arsenal of therapeutic action becomes.

Overall, the training strategies and potential benefits described above provide examples of how being able to rate psychotherapy sessions has the potential to improve trainee abilities as a therapist and researcher, and may accelerate the learning process.

Video Training Strategies for Advanced Therapists

For the experienced therapist, the process of watching and rating recorded sessions is also extremely valuable. The experienced therapist is obviously at a different developmental stage than the trainee thus the emphasis is less focused on the "how to" of psychotherapy, and more directed toward the continued cultivation and development of therapeutic skills. Using a sports analogy, less focus is paid on learning how to swing a golf club and more emphasis is placed on improving the swing. This process provides an invaluable way to continue improving one's therapeutic skills and also provides the therapist with a chance to review key moments in therapy.

We have found that experienced therapists benefit from this model of training if they integrate a 2-hr block of time on a weekly basis into their schedule. The ideal situation is to do this work with a group of therapists, but ATOStrainer is designed for the solo therapist, as well. One reason we encourage a group structure is that by sharing perspectives, others see what might have not otherwise been noticed.

Typically, groups begin by reviewing APA DVDs and scoring them, but as the group matures, participants eventually decide to video and score their own sessions for peer and/or individual supervision.

How to structure the time to review and score video is fairly simple: (1) Review how your work was impacted by the previous week of watching and scoring video; (2) Pull out a clipboard that has a copy of the ATOS and paper for taking notes; (3) Select a video and watch it in 10-min segments (sometimes people choose to review the entire video first and then go back and do 10-min segments); (4) Decide what adaptive feeling the patient needed to access or did access in each 10-min segment; (5) Score the ATOS scale; (6) Discuss your reasons for the ATOS scores and write them down; (7) Input scores into ATOStrainer; (8) Review results and reflect.

The practice of consciously watching master sessions and then scoring the sessions on a weekly basis, sharpens the therapists' eye, so they can more accurately and consistently note how their patients blocks to accessing key therapeutic markers (e.g., patient ability to: access to adaptive feeling, reality testing, and gain insight into their difficulties). Exposure to master therapist personalities and interventions remind therapist that there is no "one way" to work effectively with patients, because there are a variety of approaches and therapist styles that equally produce positive results for patients. Reviewing the video can be inspirational and motivates therapist to practice new skills and learn to trust their intuition and natural style.

It is our experience that the process of carefully analyzing psychotherapy sessions increases empathy for the patient's struggle. For example, from an STDP perspective, it helps the therapist to clearly see how maladaptive defenses and high anxiety is preventing their patient from accessing adaptive affect, which helps them to experience change. That is, we have learned over and over again that when patients feel true sadness and allow themselves to cry this allows them to experience deep relief.

Use of videotape process ratings not only increases therapists' capacity to clearly see how the patient is getting in the way of his or her own healing, but also teaches therapists helpful interventions to encourage change. We have discovered that when we stray from our practice of reviewing and coding video, we are unable to offer our patients and ourselves the gift of working at our highest level thus leaving ourselves feeling ungrounded, unsatisfied, and our sessions not as useful and gratifying. On many occasions we have gone from studying a recorded session to an actual session, and been able to catch important nuances that we might have otherwise missed, like our patients noting physical discomfort that if explored might have uncovered adaptive feeling.

After a time of videotape reviewing and process rating, therapists become so accustomed to challenging themselves and working at a highly conscious level that they notice subtle differences occur in their work during the weeks they have not been reviewing sessions. Perhaps, they inadvertently enable a patient's defenses simply by not sustaining focus in the session or miss signs of emerging affect.

Another outcome of this type of practice is that many of us begin to videotape our own sessions and then use the ATOS to score them. Because we often review our own cases for peer supervision, we are able to show segments of our work where we are having trouble. Showing video takes a peer supervision group to a whole other level where therapists are vulnerable with one another by sharing their work and receiving feedback. The act of showing ones video immediately increases the therapist's self-esteem (if it is a healthy group of peers who know how to give critical feedback) and it allows them to learn what his or her strengths and weaknesses are and what areas they need to practice. Recently, we had a seasoned therapist new to this process show his work, we scored it with the ATOS, and discussed the results. He said that in his next sessions he was able to help his patient's experience adaptive affect with him and he felt more focused.

Another benefit of the advanced therapist showing video is that it allows for them to demonstrate different approaches they used to help their patients experience strong therapeutic change. To share these "therapeutic victories" with their colleagues can be rewarding and it creates an environment where everyone is excited to watch one another develop into a stronger therapist. One would think that this might create an environment of envy, but we found it to be just the opposite, because we learn through video that therapists have distinct personalities, unique interventions, strengths, and weaknesses. As we become familiar with one another, we discover that we all have something that works well with our patients, we just may need to cultivate our skills by practicing and receiving useful feedback. Sharing what is working in therapy is essential in order to understand our strengths and increase our overall self-esteem. The better we feel about our work, the better we'll be inside the therapy session. This model of training has been essential to helping us learn and grow as therapists.

Additionally, for the advanced therapist watching and rating sessions can be very useful in the supervision context whether it is with colleagues or beginning therapists. Specifically, theoretical clinical evaluations can be linked directly to observed patient and therapist behavior. To be able to pin down often rather abstract psychological constructs and terms to observable processes in therapy, we believe can reduce misunderstandings of how we use these terms. For example, “fear of closeness” can be perceived as central problem for a patient. By knowing a psychotherapy process measure such as the ATOS instrument well, you are able to demonstrate observationally to beginning therapists or colleagues when this phenomena presents itself on the video, evaluate how this was handled in the therapy and this knowledge can further guide future therapy.

Problems and Pitfalls of Video Coding

As we noted above, learning to do therapy is accomplished through a master-apprentice model. We acknowledge that video coding alone cannot replace face-to-face supervision with a senior therapist. However, we believe that all supervision should involve the supervisor and trainee watching tapes together and then discussing clinical issues because it would provide for a higher quality learning experience. Not only should the trainee show recording, but also the supervisor as well. This could serve as an important training strategy as it would help reduce trainee anxiety and provide effective modeling for the trainer. It will also provide an opportunity for both the trainee and supervisor to practice sharing and receiving constructive feedback regarding their clinical work. Additionally, the supervisor-supervisee can discuss in greater detail a particular interaction that occurred in the therapeutic encounter whereas a Website may provide only a limited specific explanation.

Also, it is not uncommon for the supervisor and the supervisee to have disagreements on how they both view the therapy encounter unfolding. These disagreements and the resolution of them are an intricate component of supervision and training that cannot be accounted for on www.ATOSTrainer.com. With ATOS Trainer, the trainee has no one to consult when, for example, he or she disagrees with the score provided for Insight in a particular segment, or when the rater disagrees with the core affect chosen. In the future, we may add an interactive comments section where raters can leave comments regarding the ratings and have expert raters reply to them.

Additionally, video coding only captures certain elements of the therapeutic encounter and the supervision process. Training in psychotherapy extends beyond the therapists’ ability to implement effective interventions and evaluating the patient’s reaction to those interventions. The capacity for self-reflection and self-growth, and the ability to examine one’s own biases, values, and countertransferences all play key roles in the development of a therapist’s clinical ability and ultimately in a therapist’s professional identity. These key areas are ones that are commonly tackled during the supervision encounter and are (a) not the direct focus of video coding and (b) cannot be dealt with specifically by using www.ATOSTrainer.com.

Another limitation of learning to do therapy based on video coding is that some individuals may have an aversion to the use of, and reliance on, technology for training. Supervisors, trainees, and

advanced clinicians may not want to deal with the technological challenges that may come with video coding and some may feel that this type of supervision takes away from the “human” way of conducting supervision. Though this is an important limitation we have done our best with this article and with www.ATOSTrainer.com to make the process of video coding as user friendly and simple as possible. Additionally, the ratings provided on the Website are in-depth and thorough, providing the most realistic learning experience possible.

Most of the limitations presented above are only problematic if you use video coding as your exclusive platform for supervision. However, in our view, we do not see video coding as a replacement for face-to-face supervision but rather a crucial extension and enhancement of the supervision and training process. The ATOS Trainer provides an additional opportunity for training therapists to practice and develop their skills without having to be physically present with their supervisors; that is, the developmental experience, for example, can happen individually from remote locations at one’s own pace.

A good way to summarize these components is that video coding allows for trainees to develop an understanding on the “how to” of therapy; the structured use of therapeutic technique. Face-to-face supervision focuses generally on case conceptualization and professional development. Both can be done in parallel and we believe each helps the trainee become a better clinician.

Future Directions: Research and Clinical Implications

Gaps in Research Knowledge

We need to conduct research in order to explore the overall value of rating recorded sessions of psychotherapy: What is its impact on therapeutic effectiveness? Does it turn graduate students into “better” therapists? How many hours of rating is sufficient to “pass” a course or clinical practicum? Is it enough to become reliable based on ICC scores? These are questions that need to be tested empirically.

If therapists are using the ATOS scale to score psychotherapy sessions, another research question could address the quality of their therapy sessions that follow the ratings of sessions. Do therapists tend to have “better” sessions if they rate previous sessions in comparison to those therapists that do not? What objectives define “better” therapy? Can the ATOS scale scores actually determine what identifies superior therapy? Also, future research could examine whether this type of training can reduce a sense of being overwhelmed when starting as a new psychotherapist and possibly prevent burnout compared to those who do not engage in such training. These and other important questions will need to be tested empirically in the future.

Airline Pilot Standards: Incorporating Video in Graduate Training

Airline pilots are required to log in hundreds of hours of practice in a “test cockpit” before being entrusted to fly a commercial plane. However, therapist-in-training are asked to work with patients at internship sites sometimes without observing a single

psychotherapy session with a master therapist. Most students do not have access to an interactive video training and those who do receive it tend to come across it based on the esoteric research interests of their professors, on their own initiative, or just serendipitously. This is unfortunate as video training may enhance learning for graduate students, just as it does for surgical residents to witness master surgeons in the operating room, and airline pilots to practice in the test cockpit.

Before a pilot can fly actual flights, he or she must complete a certain number of simulated flight hours. These simulated flights provide the pilot an opportunity to learn about the intricacies of flying in different weather conditions, to different destinations, and during different times of the day. They gain valuable training experience that better prepares them for the real flight. Just as the pilots who will be entrusted with the lives of their passengers, graduate students will be entrusted with the lives of their patients. As such, having graduate students complete a certain number of hours of, "simulated flight conditions" like which airplane pilots face would be immensely valuable (McCullough-Vaillant, 1997, p. 525). However, instead of different weather conditions or destinations, graduate students will gain experience watching and recording tapes of different therapeutic conditions, different diagnoses, and different therapists' styles.

Watching and rating recorded videotaped sessions may be valuable to the therapist-in-training; it might be incorporated more formally and comprehensively into all graduate training of mental health professionals. For example, several universities in Norway have incorporated analysis of videotapes of therapy as courses for graduate credit, or offered as extra learning experiences. Students at the Norwegian University for Science and Technology have, for instance, been offered a choice between writing a paper, and training up to reliability on ATOS, and then coding 10 sessions of actual therapy, as part of their masters program in psychology. Reliability training was done in groups, but rating was done in pairs, allowing the students to discuss their ratings and experiences. This coding forms part of the data material in Kallestad et al. (in press). Students' anecdotal reports are very positive and over 100 students to date have taken courses for no credit just for further learning (e.g., Schanche, Hoffart, Neilson, & McCullough, in press). Students that engaged in the rating process described feeling more competent and secure in their therapy than their peers who have not had this experience.

Additionally, Pal Ulvenes (third author) and Lene Berggraf (fourth author) were members of the PROCMAP (Process Mapping Project) and during the spring and summer of 2009, 50 psychology students in the clinical psychology program at the Norwegian University of Science and Technology and the University of Oslo were assigned as Research Assistants to perform rating of the extensive video material (1,500 therapy sessions) of clinical therapy from a RCT study (Svartberg, Stiles, & Seltzer, 2004). Raters were given training for the two process instruments; 3 days of training and practice for the ATOS and 2 days of training and practice for the *Psychotherapy Process Q-Set*. The seminars provided an introduction to each instrument and good opportunity to practice using the instruments. In addition, to learn how to analyze psychotherapy, the students were given the opportunity to write a master thesis on existing data and about 12 students are in the process of writing their thesis.

The students who chose to participate on the PROCMAP project wanted to get the opportunity of observing therapists working within two different theoretical orientations of psychotherapy, to be able to see different types of patients, to learn theory, and by combining this with learning, how to analyze what happens in a therapeutic session. Most of the students reported this to be very valuable while working with their own patients, increasing their understanding of therapeutic work and processes, knowing where the patient "is" on various common factors, and to have a sense of knowing what "good" versus "bad" therapy is.

This experience highlights the fact that it is seldom students have the ability to observe experienced therapists doing their therapeutic work. Universities seem to not have enough resources to provide this type of training. This is unfortunate as based on feedback we have received from students engaging in these projects, most of them would find this useful in their learning process. In addition, this might be a way of giving students more clinical experience, without it being cost demanding.

Our goal is to provide these types of learning opportunities on www.ATOStrainer.com. Students can purchase APA DVDs, log into the site, and then practice rating each 10-min segment until they fall within the acceptable reliability range. Students will gain exposure to patients with different types of: (a) *psychiatric disorders*, for example, major depression, generalized anxiety, personality disorders, and so forth; (b) *emotional difficulties*, for example, intense sadness or grief, explosive anger, inability to experience self-compassion, and so forth; and (c) *therapeutic situations*, for example, treatment resistant patients, overly compliant patients, therapeutic ruptures, and so forth.

Ultimately, this Website gives students the opportunity to log in hours of "simulated flight hours" in the "test cockpit" before entering their first session with a patient. Once beginning to provide treatment the "trained" ATOS rater, can then evaluate his or her own sessions. This is an opportunity that may be immensely powerful, and an opportunity that should be considered for graduate training programs sooner rather than later.

Conclusions

This article highlights how video rating can be a rich and enhancing learning experience for both the student and seasoned therapist providing them the opportunity to develop and continually practice their therapeutic skills. With the step-by-step procedures outlined in this article, and with www.ATOStrainer.com, there now exists a site where this learning can be actualized. It is our hope that this article provides readers with the information needed and the inspiration to go to our site and try it out!

References

- Bhatia, M., Gill Rodriguez, M., Fowler, D., Godin, J. E. G., Drapeau, M., & McCullough, L. (2009). Desensitization of conflicted feelings: Using the ATOS to measure early change in a single-case affect phobia therapy treatment. *Archives of Psychiatry and Psychotherapy, 1*, 31-38.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: Wiley.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613-619.
- Kallestad, H., Valen, J., Svartberg, M., McCullough, L., Stiles, T. (in

- press). Insight as a mechanism of change in short term psychodynamic psychotherapy. *Psychotherapy Research*.
- McCullough, L. (2003). Videotaping and rating your own sessions: Two valuable ways to improve the quality of therapy. Retrieved from <http://www.affectphobia.org/video.html>. Retrieved July 12, 2010.
- McCullough, L., Kuhn, N., Andrews, S., Hatch, D., Valen, J., & Osimo, F. (2002). The reliability of the achievement of therapeutic objectives scale: Five studies, *Journal of Brief Psychotherapy*, 2, 1–14.
- McCullough, L., Larsen, A. E., Schanche, E., Andrews, S., & Kuhn, N. (2003a). Achievement of Therapeutic Objectives Scale: Representing well-established common factors in psychotherapy. Retrieved from <http://www.affectphobia.org/atos.html>. Retrieved July 17, 2010.
- McCullough, L., Kuhn, N., Andrews, S., Kaplan, A., Wolf, J., & Hurley, C. L. (2003b). *Treating affect phobia: A manual for short-term dynamic psychotherapy*. New York, NY: Guilford Press.
- McCullough-Vaillant, L. (1997). *Changing character: Short-term anxiety regulating psychotherapy for restructuring defenses, affects, and attachments*. New York, NY: Basic Books.
- Schanche, E., Hoffart, A., Neilson, G., & McCullough, L. (in press). The effects of varying amounts of training on the reliability of graduate students' rating of complex clinical phenomena. *Scandinavian Journal of Psychotherapy*.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Svartberg, M., Stiles, T., & Seltzer, M. (2004). A randomized controlled trial of the effectiveness of short term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *Journal of Consulting and Clinical Psychology*, 161, 810–817.
- Valen, J. (in press). The Achievement of Therapeutic Objective Scale: Inter-rater reliability and sensitivity to change in short-term dynamic psychotherapy and cognitive therapy. *Psychological Assessment*.

Received January 20, 2011

Accepted January 22, 2011 ■

Online First Publication

APA-published journal articles are now available Online First in the PsycARTICLES database. Electronic versions of journal articles will be accessible prior to the print publication, expediting access to the latest peer-reviewed research.

All PsycARTICLES institutional customers, individual APA PsycNET® database package subscribers, and individual journal subscribers may now search these records as an added benefit. Online First Publication (OFP) records can be released within as little as 30 days of acceptance and transfer into production, and are marked to indicate the posting status, allowing researchers to quickly and easily discover the latest literature. OFP articles will be the version of record; the articles have gone through the full production cycle except for assignment to an issue and pagination. After a journal issue's print publication, OFP records will be replaced with the final published article to reflect the final status and bibliographic information.