AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychotherapy

© 2022 American Psychological Association
ISSN: 0033-3204

2023, Vol. 60, No. 2, 149–158
https://doi.org/10.1037/pst0000460

You Never Know What You Are Going to Get: Large-Scale Assessment of Therapists' Supportive Counseling Skill Use

Xinyao Zhang[1], Michael Tanana[2], Lauren Weitzman[1], Shrikanth Narayanan[3], David Atkins[4], and Zac Imel[1]

[1] Department of Educational Psychology, University of Utah
[2] Social Research Institute, College of Social Work, University of Utah
[3] Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California
[4] Department of Psychiatry and Behavioral Sciences, University of Washington

Supportive counseling skills like empathy and active listening are critical ingredients of all psychotherapies, but most research relies on client or therapist reports of the treatment process. This study utilized machine-learning models trained to evaluate counseling skills to evaluate supportive skill use in 3,917 session recordings. We analyzed overall skill use and variation in practice patterns using a series of mixed effects models. On average, therapists scored moderately high on observer-rated empathy (i.e., 3.8 out of 5), 3.3% of the therapists' utterances in a session were open questions, and 12.9% of their utterances were reflections. However, there were substantial differences in skill use across therapists as well as across clients within-therapist caseloads. These findings highlight the substantial variability in the process of counseling that clients may experience when they access psychotherapy. We discuss findings in the context of both the need for therapists to be responsive and flexible with their clients, but also potential costs related to the lack of a more uniform experience of care.

Clinical Impact Statement

Question: We conducted a large-scale observational study to address the knowledge gap about clinicians' use of basic counseling skills, such as empathy and active listening skills. ***Findings:*** We found that clinicians varied substantially in their use of basic counseling skills across clinicians as well as within their caseloads. ***Meaning:*** This variation suggests that quality improvement efforts in behavioral healthcare might focus not only on the implementation of specific treatments but also ensuring the use of basic counseling skills. ***Next Steps:*** Future studies will continue to explore the impact of this variation on psychotherapy outcomes.

Keywords: supportive skills, machine learning, therapist difference, linear mixed modeling, naturalistic study

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

There are a number of forms of effective psychotherapy (Wampold & Imel, 2015). Common factors or treatment elements common to established treatments are responsible for a significant portion of these treatment effects (Alldredge et al., 2021; Constantino et al., 2018; Horvath et al., 2011; Irving et al., 2004; Probst et al., 2019). Consistent with this evidence, a meta-analysis of 31 studies indicated that nondirective supportive counseling could achieve almost 75% of the total pre–post treatment effect size of active psychotherapies (Cuijpers et al., 2012). Ensuring that therapists are prepared to offer clients these foundational skills such as empathy and active listening is critical to providing a base of quality care to clients.

Supportive counseling is characterized by key interpersonal techniques such as active listening, reflections, encouraging, and helping clients explore their experiences and emotions, instead of

This article was published Online First October 27, 2022.

Xinyao Zhang https://orcid.org/0000-0001-6227-3205

Michael Tanana, Shrikanth Narayanan, David Atkins, and Zac Imel share minority equity interests in Lyssn.io, a technology company focused on developing technologies to evaluate the contents of psychotherapy. Shrikanth Narayanan is also the chief scientist and cofounder with equity stake in Behavioral Signals, a technology company focused on creating technologies for emotional and behavioral machine intelligence. This study is funded by the National Institute on Alcohol Abuse and Alcoholism (R01AA018673-06A1) awarded to David Atkins.

Xinyao Zhang played lead role in conceptualization, formal analysis and writing of original draft and equal role in methodology. Michael Tanana played supporting role in conceptualization, formal analysis, methodology and software and equal role in data curation. Lauren Weitzman played lead role in data curation, supporting role in resources and equal role in project administration and writing of review and editing. Shrikanth Narayanan played supporting role in supervision and equal role in funding acquisition, project administration, resources and software. David Atkins played supporting role in conceptualization and supervision and equal role in funding acquisition, project administration and resources. Zac Imel played lead role in supervision and writing of review and editing, supporting role in conceptualization and formal analysis and equal role in funding acquisition, methodology, project administration and resources.

Correspondence concerning this article should be addressed to Xinyao Zhang, Department of Educational Psychology, University of Utah, 1721 Campus Center Drive, Salt Lake City, UT 84112, United States. Email: xzhang@sa.utah.edu

giving advice or solving problems (see Areán et al., 2010). Taking various forms, supportive counseling is often operationalized in clinical trials as an attempt to provide a supportive, attentive therapist that might be helpful for a client, but not to provide any specific psychological treatment (e.g., cognitive–behavioral therapy [CBT], interpersonal psychotherapy [IPT]). Despite the long history of supportive counseling (Conte, 1994; Rockland, 1993; Winston et al., 1986) and evidence that simply listening actively in a supportive, empathic manner can be powerful (Cuijpers et al., 2012), the extent to which therapists regularly use the basic skills that are thought to be common in clinical practice remains unclear.

Assessment of Differences in Therapist Skill Use

Driven by concern that providers rarely use specific evidence-based treatments, work evaluating the skills utilized by therapists in community settings has focused largely on the dissemination and implementation of specific evidence-based treatments (McHugh & Barlow, 2010). Even after training, many providers fail to sustain the use of the treatments they were trained to perform, drifting away from the core elements of the treatments over time (Schwalbe et al., 2014; Wiltsey Stirman et al., 2013). For example, in the largest study of its kind, Creed et al. (2016) found in a study of over 300 therapists (and almost 1,000 observer-rated sessions), that therapist endorsement of a CBT orientation was not related to use of CBT skills. Therapists' adherence to a specific treatment is subject to various sources of influence during the therapy session, including both client factors (Boswell et al., 2013; Imel et al., 2011; Snippe et al., 2019) and therapist factors (Snippe et al., 2019; Wiltsey Stirman et al., 2015). Therapists can account for about 13%–40% of the total variance in adherence and competence to an intervention depending on the setting and type of intervention (Boswell et al., 2013; Imel et al., 2011, 2014). These studies also show that the client that a therapist worked with had a similarly large effect on use of various interventions, accounting for 14%–32% of the total variance in therapist adherence and competence. Two other studies (Dunn et al., 2016; Hallgren et al., 2018) focused on single-session motivational interviewing interventions and also demonstrated substantial variation in the skill use between therapists. Given this evidence for use of specific treatments, it is plausible that there is similar therapist variation in the use of supportive counseling skills.

One factor that might impact estimates of therapist performance in naturalistic settings is a reliance on human observers, which limits the scale of studies such that large representative samples are very difficult and slow to obtain. Specifically, therapists select which recordings to send and only a limited number can be rated. Studies that rely on therapist self-report of evidenced-based practices may be subject to social desirability, recall, or other biases. Larger studies like Creed et al. (2016), can take many years to execute given the demands of human coding. However, there are emerging tools that make it possible to assess the therapist's skill use on a larger scale.

Machine Learning-Based Evaluation of Psychotherapy

Natural language processing refers to a set of technologies used to transform unstructured text data (e.g., therapy transcripts) to more structured and useful information (see Pace et al., 2016, for discussion related to psychotherapy). At present, large-scale efforts to evaluate the quality of psychotherapy rely on human-coding efforts that require hiring, training, and supporting teams that directly observe and rate sessions. This process is slow and expensive, and thus limits the scale of coding efforts even in the most ambitious and well-funded studies (Imel et al., 2017). Modern machine-learning-based technologies can utilize human ratings of psychotherapy sessions to train models that allow automatic recognition of key treatment processes, such as session topics (Dinakar et al., 2015; Imel et al., 2015), emotions (Luo et al., 2020; Maskit, 2021; Tanana et al., 2021), motivational interviewing fidelity (Atkins et al., 2014; Can et al., 2016; Imel et al., 2019; Xiao et al., 2012), CBT (Ewbank et al., 2020; Hilbert et al., 2020), and discussions of suicide (Walsh et al., 2017). In particular, these technologies can label core aspects of supportive counseling like empathy (Gibson et al., 2016; Xiao et al., 2012, 2016; Xiao, Imel, Atkins et al., 2015; Xiao, Imel, Georgiou et al., 2015), interpersonal skills (Goldberg et al., 2021), open questions (Flemotomos et al., 2022), and reflections (Can et al., 2016); see Aafjes-van Doorn et al. (2021), for a detailed review.

One advantage of machine learning is the ability to scale up the sample size of psychotherapy studies from tens or hundreds to thousands of sessions. For example, Ewbank et al. (2020) trained a machine-learning model to recognize indicators of CBT in text-based counseling. They then utilized this model to evaluate the use of CBT in a large sample (90,934 text-based sessions). As this study relied on text-based counseling, it avoided a number of complications related to capturing session audio reliably, automated speech recognition, utterance segmentation, diarization (i.e., separation of speakers), and role assignment (i.e., determining who the therapist is). However, Flemotomos et al. (2022) described an online system that manages these practical and computation problems and facilitates direct evaluation of spoken language counseling at scale.

The Present Study

Foundational education in psychotherapy includes training in basic supportive counseling skills, such as the techniques designed to communicate empathy and facilitate therapeutic relationships. Indeed, these skills represent the necessary background of many evidence-based treatments even if they are not considered the key ingredients. However, studies of therapist skill use in community settings are primarily focused on evaluating the impact of efforts to disseminate some specific type of evidence-based treatment, rather than focusing directly on the core skills that are important across treatments. It is possible that just as with evidence-based treatments, the use of these skills is far less prevalent than stakeholders might hope. As supportive counseling can have a substantial impact on clinical symptoms (Cuijpers et al., 2012), it is then important to assess the degree to which therapists use core supportive counseling skills.

In this study, we conducted the largest study of therapist behavior during in-person counseling to date.[1] We used a machine-learning model trained on more than 2.8 million human-coded statements (see Flemotomos et al., 2022) to estimate therapists' use of several

[1] We cited a large study by Ewbank et al. (2020), but this was a text-based study. We are aware of other large-scale implementation studies (Clark, 2011; Karlin & Cross, 2014; McHugh & Barlow, 2010) that included large numbers of therapists and clients. However, these studies were limited by missing information about the within-session therapy processes, such as statement-level coding of intervention as in this study.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

basic counseling skills in a large collection of sessions. The data set was collected under an research project on motivational interviewing approved by the institutional review board (IRB_00083132). We hypothesized that therapist variability would be larger than previously found in dissemination studies (e.g., Boswell et al., 2013; Dunn et al., 2016; Hallgren et al., 2018; Imel et al., 2011), but that therapist performance would also vary across clients. As mentioned earlier, most of the previous studies were limited to particular treatment types, were limited in size, or relied on sessions obtained during treatment implementation efforts, which may have reduced variability in therapist performance due to training, selection, and supervision efforts. In Hallgren et al.'s (2018) study, for example, therapists accounted for 22% of the variability in empathy in training studies, which was most close to a naturalistic setting compared to efficacy and effectiveness studies. We expected therapist variability in our study to be close to or above this number.

Method

Data Source

After cleaning, the data set for the present study included recordings of 3,917 counseling sessions, or 2,624,298 utterances, obtained through an ongoing study of machine-learning-based observation of psychotherapy process at a large counseling center at a university in the Western United States (Flemotomos et al., 2022). The data set contained 1,011 clients and 57 therapists. Sessions were digitally recorded in an audio form. Both therapists and clients had given formal consents to the recording of the sessions. The sessions were recorded with two microphones hung from the ceiling of the therapy rooms, one omnidirectional and one directed to the therapist. The recorded sessions were processed in an automated pipeline that includes five steps: (1) voice activity detection, (2) speaker diarization, (3) speech recognition, (4) speaker role recognition, and (5) utterance segmentation. The pipeline converts audio recordings to speaker-labeled transcripts for each session and then assigns therapist skill codes at the utterance level, as well as several additional session-level tags, as described below (Flemotomos et al., 2022).

The client sample self-identified as 54.1% female, 42.0% male, and 3.9% other genders. The ethnic composition was 73.2% European Americans, 9.5% Asian Americans, 7.5% Latino/Latina Americans, 5.7% multiethnic, and 1.6% African Americans. Ages ranged from 18 to 54, with a mean of 23.4 and a median of 22. The therapist sample had similar demographic properties. For therapists, 56.1% identified as female, 26.3% identified as male, and 3.5% identified as genderqueer. In terms of ethnicity, 56.1% identified as European Americans, 8.8% Latino/Latina Americans, 5.3% Asian Americans, and 3.5% African Americans. Ages ranged from 21 to 69 years, with a mean of 32.8 and a median of 29. In the therapist sample, 18.2% self-identified as cognitive–behavioral therapy oriented, 15.5% as multicultural/feminist oriented, 12.8% as interpersonal therapy oriented, 12.2% as integrated/eclectic, 10.8% humanistically oriented, and 5.4% were psychodynamically oriented.

Measures

Therapist use of basic counseling skills was assessed with a machine-learning model trained on human ratings of psychotherapy sessions (Flemotomos et al., 2022). For this study, we selected specific items that assessed basic active listening and counseling skills from the Motivational Interviewing Skill Code 2.5 (MISC 2.5; Houck et al., 2010). In this study, we used three measures at the session level: (1) empathy, (2) count of reflections, and (3) count of open questions. Empathy is defined as the extent to which therapists understand or attempt to understand the client's perspective. The rating of empathy ranges from 1 (*low*) to 5 (*high*). A low rating is defined as the therapist showing little interest in the clients' own experiences. In contrast, high empathy is characterized by the assessed counselor showing active attentiveness to the clients' worldview.

Two types of reflective statements, simple reflection and complex reflection, were combined into one overall count of reflections. A simple reflection is defined as the therapist capturing and returning to the client something that the client has said previously in the same session, without adding meaning or emphasis to the client's statement. Simple reflection is used to merely convey understanding or facilitate conversation. An example of a simple reflection could be a therapist saying "You feel sad; and you want to hide away" after the client said, "I feel sad and just want to hide." A complex reflection enriches what the client conveys by adding significant meaning or emphasis to the client's statement. For example, a therapist may say: "It sounds like you feel angry when your colleague did that to you" after the client shared that her colleague talked behind her back.

Open questions are questions that leave latitude for responses, instead of the questions that can often be answered with simple responses such as yes or no. An open question oftentimes begins with wh-words and one example could be "What does it mean when you said I just feel it's impossible?" (Houck et al., 2010).

Because the system predicts reflection and open questions at the utterance level, we calculated the frequency of each measure in a session to obtain the session-level counts. Empathy is scored at the session level and does not need further transformation. The performance of the system was evaluated by accuracy and F_1 score, with the latter being the harmonic mean of precision and recall. The accuracy for predicting high or low empathy was .851. The F_1 score was .342. The F_1 scores for reflections and open questions were .612 and .825, respectively (Flemotomos et al., 2022).

Analysis

Statistical Modeling

We used generalized linear mixed modeling (GLMM) to examine differences in use of the basic skills highlighted above. GLMM is a generalized version of linear mixed modeling (LMM) or multilevel modeling. LMM fits the nested structure of counseling (Baldwin et al., 2007), where one therapist often provides care to multiple clients, and clients attend multiple sessions. This nested structure violates the independence assumption of commonly used statistical methods such as analysis of variance and multiple linear regressions. Compared to LMM, GLMM is more flexible with the choice of model distributions and therefore can model assorted types of data, such as time length (continuous), dropout (binary), or the frequency of using a particular intervention (count). In this study, we constructed three-level empty GLMMs with sessions (Level 1)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

nested in clients (Level 2) nested in therapists (Level 3). For continuous variables (i.e., empathy), we used a normal distribution. With normal distributions, GLMM is equivalent to LMM. In particular, empathy was modeled with robust-error LMM.[2]

$$y_{ijk} \sim N(\mu_{ijk}, \varepsilon_{ijk}), \tag{1}$$

$$\text{Level 1: } \mu_{ijk} = \beta_{0jk}, \tag{2}$$

$$\text{Level 2: } \beta_{0jk} = \gamma_{00k} + u_{0jk}, \tag{3}$$

$$\text{Level 3: } \gamma_{00k} = \delta_{000} + v_{00k}, \tag{4}$$

Y_{ijk} is the raw empathy score for client j and therapist k in session i, μ_{ijk} is the estimated mean of the empathy scores and ε_{ijk} is the residual uncounted by the model. β_{0jk} is the mean score across all sessions of client j, treated by therapist k, representing the mean empathy level with a particular client or client–therapist dyad. γ_{00k} is the mean score across all clients within therapist k, representing the mean empathy level of a particular therapist. δ_{000} is the grand mean, representing the average empathy level of the entire sample. u_{0jk} and v_{00k} are random-effect variables and assumed to be multivariate normal centered at 0 and have a covariance matrix with τ_j^2 and τ_k^2 on the diagonal. For count variables (i.e., reflection and open questions), we used a negative binomial distribution with offset terms and a log link[3] to model the data.

$$Y_{ijk} \sim \text{NB}(\mu_{ijk}, \alpha), \tag{5}$$

$$\text{Level 1: } \ln(\mu_{ijk}) = \beta_{0jk} + \ln(N_{ijk}), \tag{6}$$

$$\text{Level 2: } \beta_{0jk} = \gamma_{00k} + u_{0jk}, \tag{7}$$

$$\text{Level 3: } \gamma_{00k} = \delta_{000} + v_{00k}, \tag{8}$$

α is a dispersion parameter accounting for the over- or underdispersion of the data. $\ln(N_{ijk})$ is an offset term accounting for the varied session lengths that may influence the estimation of count variables (i.e., a longer session may likely have more reflections than a shorter session). Here, N_{ijk} is the total number of therapist's talk turns for therapist k in session i with client j. The addition of an offset term under a log link essentially converted the estimation of the mean counts to the estimation of the mean proportions. For example, an estimated δ_{000} of -1.39 for reflection means that, on average, 25% of the therapists' utterances are reflections (the natural exponent e to the power of -1.39 is about .25).

For each of the three variables, client and therapist effects were assessed using intraclass correlation (ICC).

$$\text{ICC}_j = \frac{\tau_j^2}{\tau_j^2 + \tau_k^2 + \sigma_\varepsilon^2}, \tag{9}$$

$$\text{ICC}_k = \frac{\tau_k^2}{\tau_j^2 + \tau_k^2 + \sigma_\varepsilon^2}, \tag{10}$$

where ICC_j represents the client effect and ICC_k represents the therapist effect. τ_k^2 is the between-therapist variance, τ_j^2 is the within-therapist variance, and $\sigma_\varepsilon^2 = \text{var}(\varepsilon_{ijk})$ is the variance of the residuals. σ_ε^2 was estimated by the summation of distribution-specific variance and additive overdispersion (Nakagawa & Schielzeth, 2017). ICC_k can be interpreted as the extent to which therapists vary among each other in their use of the skill assessed. A high ICC_k implies a strong therapist influence on that particular skill. For example, an ICC_k of .13 for empathy means that 13% of the variability in the empathy scores is due to some differences between therapists. Likewise, ICC_j can be interpreted as the extent to which therapists vary in their use of a skill among the clients within their caseloads.

All analyses were conducted in *R* (R Core Team, 2020). GLMMs were implemented using the *lme4* package (Bates et al., 2015). Robust-error LMM was implemented with the *robustlmm* package (Koller, 2016). ICCs were calculated using the *performance* package (Lüdecke et al., 2021).

Data Reduction

We conducted a series of data cleaning steps before modeling the data. First, as these data were collected during regular clinical practice, we expected to encounter a number of recordings that were problematic (i.e., accidental recordings, microphone issues, etc.). We inspected the density distributions of the sessions based on session length, number of utterances, and therapist talk time ratio. We set cut-off points to remove long tails in the density distribution plots, which were considered outlier sessions. In particular, we removed any recording shorter than 20 min or longer than 70 min, or with less than 110 or more than 1,300 utterances. These sessions constituted 4.3% of the total sample size. In addition, there were cases where automatic separation of speakers failed—resulting in sessions where speaker labels were almost uniformly therapist or client. To address this problem, we removed sessions where therapist talk time made up either less than 10% or more than 75% of the session. The cleaned data set contained 3,917 sessions, 1,011 clients, and 57 therapists, which was 91.8% of the total collected data ($N = 4{,}269$).

Second, we censored skill labels from brief therapist statements. As described in Flemotomos et al. (2022), the automatic system not only transcribes the session and labels counseling skills, but it also uses a sentence parser to determine when a therapist talk turn starts and stops (i.e., where a label should be applied). This is a challenging problem in spoken dialogue with interruptions, disfluencies, back-channels, within speaker pauses, and so forth, and thus introduces another source of error—that is, to what text should the system provide a counseling skill label (indeed, this is a source of error almost never considered in psychotherapy research). As we inspected the raw transcript data, we found a higher preponderance of brief talk turns in auto-generated transcripts compared to human-generated transcripts. Specifically, in some cases, the system appeared to be too sensitive in segmenting talk turns. As a result, in some instances, the automatic system has very limited linguistic information to accurately predict a label. To reduce noise from brief talk turns when the system has limited linguistic information,

[2] We first tested regular LMM on the empathy score and found that the estimated model was heteroscedastic, which violated the normal distribution assumption. We then switched to robust-error LMM to obtain a more accurate estimation of the variances.

[3] We also tried several other methods, including Poisson GLMM, negative binomial GLMM with session length as covariate, four-level logistic GLMM, and four-level Bayesian logistic GLMM. We chose negative binomial GLMM with offset terms because it was the most parsimonious model that fits the data type and distribution.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

we removed the skill labels of those therapist talk turns with less than eight words prior to statistical modeling.[4]

Results

The descriptive statistics of the three assessed variables after data cleaning are shown in Table 1. The therapists' average empathy score was 3.86. On average, there were 36 (12.9%) coded reflections and nine open questions (3.3%) per session. As an illustration, we have provided sample text of coded statements in two sessions in Figures 1 and 2.

Very brief statements (less than 8 words) represented 61.7% of the total talk turns, but only 25.1% ($N = 2{,}713{,}071$) of the total words. The percentage of open questions before removing the brief talk turns was 7.3% ($N = 82{,}137$), and reflections were 34.7% ($N = 390{,}921$). If brief responses are entirely removed from the analysis (i.e., not considered in the count of messages), the percentages remain relatively stable with 8.2% in open questions, and 33.0% in reflections. If the brief talk turns labels are treated as missing but the messages are still included in session counts, the relative frequency of reflections (3.1%, $N = 35{,}252$) and open questions (12.6%, $N = 142{,}575$) and open questions are substantially reduced. This indicates that therapists on average spent 3.1% of all their speech using reflections, and 12.6% using open questions.

As indexed by the ICC, therapist variation in the use of each skill was relatively large. Therapists account for 16.9% (ICC_k) of the variability in empathy ratings, 23.4% of the variability in open questions, and 25.7% of the variability in reflections. As mentioned earlier, these numbers imply to what extent therapists differed in empathy ratings. For example, 16.9% of the variability in empathy ratings can be explained by the differences in therapists per se.

Differences in average therapist skill use across clients were similarly large, if not somewhat larger than therapist differences. The client ICC_j for empathy was 38.1%, which was about twice the therapist effect of the same rating. The client ICC_j of the counts of open questions and reflections were 27.0% and 31.8%, respectively. This suggests that the differences in client–therapist dyads had a much stronger influence on empathy ratings than therapists. On the other hand, the frequency of the use of open question and reflection was equally affected by both client–therapist dyads and by therapists. The mean and variance estimates of the three variables are shown in Table 2.

Discussion

We used generalized linear mixed effects modeling to examine therapists' use of basic counseling skills in a machine-learning augmented data set from a counseling center at a large university in the Western United States. In line with prior findings (Boswell et al., 2013; Imel et al., 2011), our study demonstrated that therapists varied among each other in their use of basic counseling skills. In addition, therapists also varied in their use of the basic skills in treating different clients. Contrary to our hypothesis, the extent to which therapists and clients varied among each other in skill use did not differ considerably from prior implementation and dissemination studies.

For the empathy rating, the therapist ICC was much smaller than the client ICC (16.9% vs. 38.1%), meaning that therapists' empathy differed more within their caseloads than it did between their caseloads. This pattern is similar to prior work on therapist differences in adherence to specific evidence-based interventions (e.g., Boswell et al., 2013; Imel et al., 2011) in that clients were a larger source of variance in adherence than therapists. However, differences between- and within-therapist caseloads were comparable and large for both open questions and reflections.

Two hypotheses may be made on why therapists differed in the use of basic skills across different clients in their caseload. First, particular clients may make it more difficult for therapists to express empathy in general. Alternatively, a particular client might be a poor fit for a therapist such that it is more difficult for the therapist to express empathy with that specific client. This phenomenon could be due to countertransference (Peabody & Gelso, 1982), or some clients might be generally less expressive, making it harder to understand their experience (Zaki et al., 2008). The second possibility is that therapists are adapting to clients' characteristics or needs at the moment. For example, one client may prefer a more didactic, solution-focused style while another client may prefer an exploratory, flexible style that utilizes reflections and questions extensively (Beutler et al., 1991). Indeed, rigidity to a specific treatment approach can be detrimental to outcomes (Castonguay et al., 2010), while therapists' flexibility or responsiveness (Constantino et al., 2020; Stiles & Horvath, 2017; Watson & Wiseman, 2021) is linked to positive process and outcomes (Katz et al., 2019, 2021; Li et al., 2020; Owen & Hilsenroth, 2014). In this situation, it is a therapist's skill in itself to be able to flexibly use these nonspecific skills depending on clients.

However, our study showed that empathy had even greater between-client variability (38.1%) than the use of reflection or open questions (27.0%–31.8%). If the observed between-client variability was mainly due to therapist flexibility, it should be smaller with empathy. Empathy is broadly considered to be a core common factor and as such, one might expect expression of empathy to be relatively less variable in skilled therapists. Therefore, it is not likely that therapist flexibility alone explains the observed between-client variability.

Future work should focus on assessing between-client variability in association with treatment outcomes. Hypothetically, the two situations (i.e., therapist flexibility vs. difficulty using specific skills)

Table 1
Descriptive Data of Supportive Counseling Skills

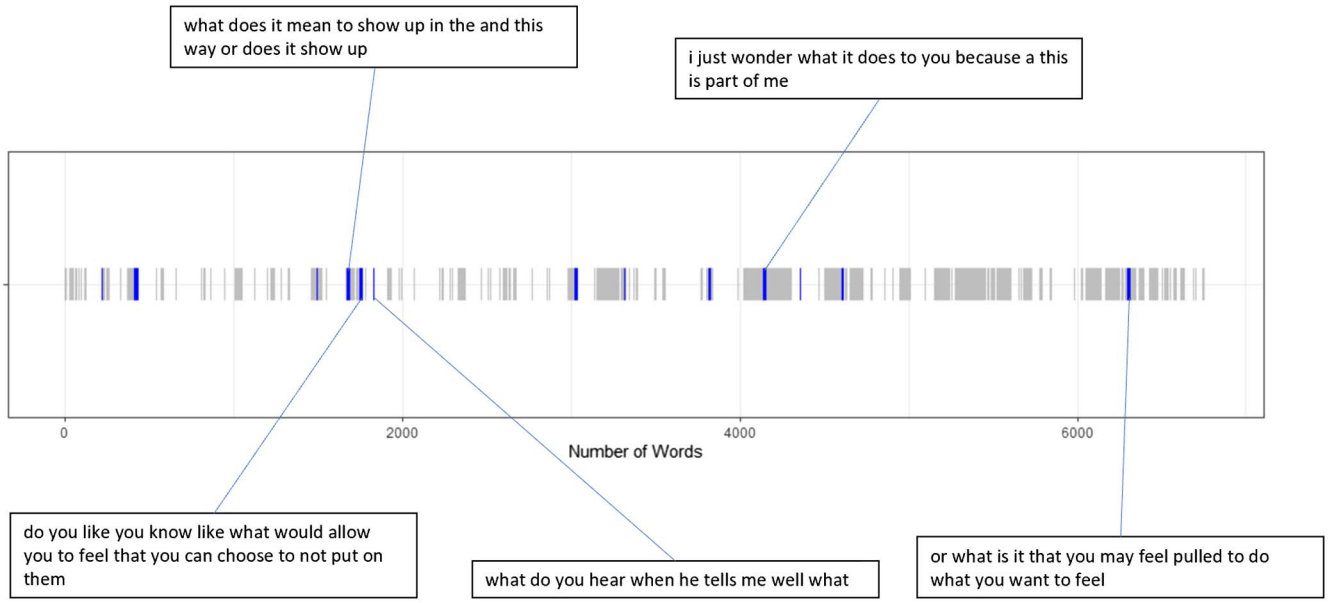
Variable	*M* (*SD*)	*Mdn*
Empathy	3.86 (0.39)	3.91
Reflection (counts)	36.44 (16.46)	34
Open question (counts)	9.16 (5.96)	8

[4] The decision to remove brief talk turns will necessarily lead to our results providing a lower bound estimate of skill use (i.e., it will remove some valid instances of therapist skills). To examine the impact of removing brief statements on statistical tests, we examined model results by progressively removing 0, 1, 2, … , and 7-word statements. Therapist's and client's ICCs were relatively stable. However, as expected, the overall frequency of counseling skills within a session decreased as we removed short utterances. Accordingly, the overall rate of skill use in these sessions should be interpreted as a lower bound of counselor skill use in session. However, removing short statements did not substantially impact estimates of differences in skill use across therapists or clients.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 1

Therapists' Use of Open Questions in an Example Session

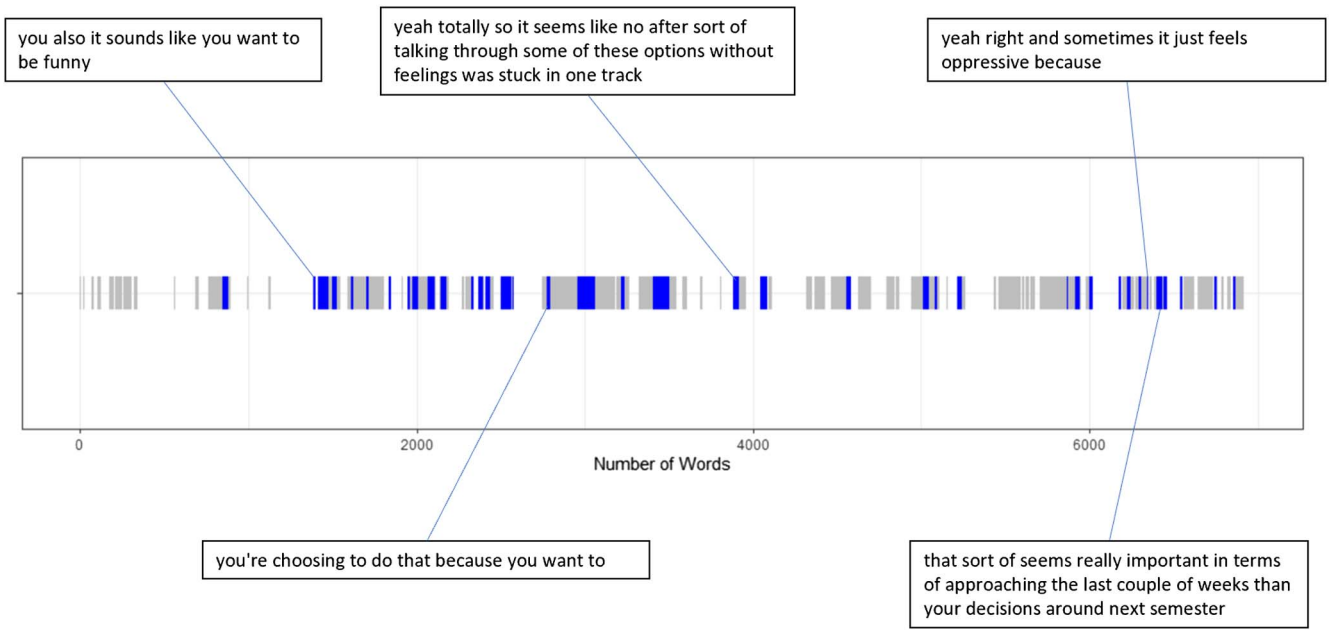


Note. The blue bars represent the therapists' utterances coded as open questions. The gray bars are therapists' utterances other than open questions. The blank areas are clients' utterances. The length of each bar is proportional to the number of words in the utterance. See the online article for the color version of this figure.

would show opposite effects. If the between-client variability positively predicts therapy processes and outcomes, it would indicate that therapists are varying skill use appropriately with different clients—actively adapting to clients' needs. If the variability negatively predicts therapy processes and/or outcomes, it indicates that there is something detrimental in therapists varying their skill use, thus pointing to the situations where differences in the use of particular skills are not helpful.

Figure 2

Therapists' Use of Reflections in an Example Session



Note. The blue bars represent the therapists' utterances coded as reflections. The gray bars are therapists' utterances other than reflections. The blank areas are clients' utterances. The length of each bar is proportional to the number of words in the utterance. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Estimates of Three-Level Models of Supportive Counseling Skills

Variable	Fixed effect (intercept)	Between-therapist variance (τ_k^2)	Within-therapist variance (τ_j^2)	Residual (σ_ε^2)	Client ICC (ICC_j)	Therapist ICC (ICC_k)
Empathy	3.829	0.023	0.052	0.061	38.1%	16.9%
Open question	0.03[a]	0.082	0.094	0.174	27.0%	23.4%
Reflection	0.124[a]	0.031	0.039	0.053	31.8%	25.7%

Note. ICC = intraclass correlation; GLMM = generalized linear mixed modeling; LMM = linear mixed modeling. Empathy score was modeled with robust-error LMM; open questions and reflections were modeled with negative binomial GLMM with the count of therapists' utterances in a session as an offset term.
[a] With offset terms, the estimated intercepts represent average proportions over the therapists' total utterances in a session. The presented values have been converted back to the original scale through log link.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Our findings also showed that therapists varied among each other in their use of supportive counseling skills. The extent to which therapists varied was consistent with findings from prior studies on specific treatments (Boswell et al., 2013; Imel et al., 2011). Knowing how therapists use these skills can be important to exploring differences in the use of specific treatments, as supportive skills are found broadly across training programs and foundational in many treatment modalities. One possible explanation for the observed therapist differences may be therapists' affiliation to particular theoretical modalities. For example, therapists who used CBT were more likely to offer advice than reflection compared to psychodynamic therapists (Wiser & Goldfried, 1996). Although reflection is seen across theoretical orientations, therapists may have different opinions on when and how to use reflection depending on their theoretical frameworks, thus contributing to the therapist differences observed in this study.

It may also be possible that the differences in the use of specific supportive counseling skills between therapists would potentially pose a concern about the quality of care that clients received, given a literature review showing that these nonspecific skills can result in a significant change on their own (Cuijpers et al., 2012). Nonspecific skills such as empathy (Elliott et al., 2011, 2018; Greenberg et al., 2001; Nienhuis et al., 2018) and the use of reflection and open questions (Magill et al., 2014, 2018; Pace et al., 2017) are consistently found to predict treatment processes and outcomes. How therapists use these basic skills are possible to play an important role in the quality of care that clients receive. Therefore, it may be of interest to assess how the competence and adherence to these nonspecific skills relate to treatment outcomes. This could help answer the question about whether the finding in the present study, that therapists differed among each other in the use of nonspecific skills, is a concern of quality of care.

Limitations and Future Directions

Several limitations are worth noting here. First, the trained model we used was not perfectly reliable with human raters (human raters are also not perfectly reliable with each other). The F_1 scores for open questions and reflections were acceptable, but there is a large discrepancy between accuracy and F_1 score for empathy (see Flemotomos et al., 2022). Another limitation is the cleaning process we applied to the data set. We found overcounting of some intervention codes due to very short utterances, which we removed. This method, although intuitive and simple, could potentially overcorrect the issue. However, the estimated ICCs remained relatively stable during this iterative process, which indicates that the distributions of therapist effect and client effect were not affected by this cleaning process. As mentioned, the machine-learning model we used may have been oversensitive in parsing statements, which resulted in the higher proportion of very short utterances compared to human parsing. These short utterances may in turn pose difficulties in the subsequent speech recognition due to limited phonetic information (Kanagasundaram et al., 2011). More recent techniques for short utterance issues in machine learning (e.g., Jung et al., 2019; Kanagasundaram et al., 2011) may benefit the model with a better approximation to the human parsing of statements.

It may also be of interest to investigate if process and outcome variables are associated with therapists' use of supportive counseling skills. Like empathy, it might be of interest to investigate if the associations based on machine-learning processed large data sets align to those found in meta-analyses, thus offering criterion validity of using such methods. In addition, although explicitly conceptualized in motivational interviewing (Houck et al., 2010), open questioning or reflections are also found in other theoretical orientations to facilitate treatment processes (e.g., Anvari et al., 2020; Apodaca et al., 2016). However, there is a limited number of studies that assess these basic skills in naturalistic settings, which may be an important step given the prevalence of nonspecific skills in all treatments. Given the heterogeneity of the sample in a naturalistic setting, various moderators such as therapeutic orientation can be added to further specify the differential use and effects of these basic skills on treatment outcomes.

In summary, our findings show that therapists varied in their use of supportive counseling skills, not only among each other but also among their caseloads. Using a pretrained machine-learning model (Flemotomos et al., 2022), we scaled up process–outcome research in a naturalistic setting, which had largely been based on human coding in the past and suffered from several limitations (Imel et al., 2017). Our findings replicated what was found in the prior studies regarding the therapist and the client effects in the use of particular skills (Boswell et al., 2013; Imel et al., 2011), which offered support to the potential use of automatic measures in scaling up process–outcome studies in psychotherapy research. This study provided evidence that with reliably trained machine-learning models, the data collected from these models can be useful for process–outcome analyses where researchers would often need human coding.

References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research.

Psychotherapy Research, *31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729

Alldredge, C. T., Burlingame, G. M., Yang, C., & Rosendahl, J. (2021). Alliance in group therapy: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *25*(1), 13–28. https://doi.org/10.1037/gdn0000135

Anvari, M., Hill, C. E., & Kivlighan, D. M. (2020). Therapist skills associated with client emotional expression in psychodynamic psychotherapy. *Psychotherapy Research*, *30*(7), 900–911. https://doi.org/10.1080/10503307.2019.1680901

Apodaca, T. R., Jackson, K. M., Borsari, B., Magill, M., Longabaugh, R., Mastroleo, N. R., & Barnett, N. P. (2016). Which individual therapist behaviors elicit client change talk and sustain talk in motivational interviewing? *Journal of Substance Abuse Treatment*, *61*, 60–65. https://doi.org/10.1016/j.jsat.2015.09.001

Areán, P. A., Raue, P., Mackin, R. S., Kanellopoulos, D., McCulloch, C., & Alexopoulos, G. S. (2010). Problem-solving therapy and supportive therapy in older adults with major depression and executive dysfunction. *The American Journal of Psychiatry*, *167*(11), 1391–1398. https://doi.org/10.1176/appi.ajp.2010.09091327

Atkins, D. C., Steyvers, M., Imel, Z. E., & Smyth, P. (2014). Scaling up the evaluation of psychotherapy: Evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science : IS*, *9*(1), Article 49. https://doi.org/10.1186/1748-5908-9-49

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, *75*(6), 842–852. https://doi.org/10.1037/0022-006X.75.6.842

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beutler, L. E., Clarkin, J., Crago, M., & Bergan, J. (1991). Client–therapist matching. In C. R. Snyder & D. R. Forsyth (Eds.), *Handbook of social and clinical psychology: The health perspective* (pp. 699–716). Allyn & Bacon.

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., Woods, S., & Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, *81*(3), 443–454. https://doi.org/10.1037/a0031437

Can, D., Marín, R. A., Georgiou, P. G., Imel, Z. E., Atkins, D. C., & Narayanan, S. S. (2016). "It sounds like …": A natural language processing approach to detecting counselor reflections in motivational interviewing. *Journal of Counseling Psychology*, *63*(3), 343–350. https://doi.org/10.1037/cou0000111

Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R., & Hill, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist*, *65*(1), 34–49. https://doi.org/10.1037/a0017330

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, *23*(4), 318–327. https://doi.org/10.3109/09540261.2011.606803

Constantino, M. J., Coyne, A. E., Boswell, J. F., Iles, B. R., & Vîslă, A. (2018). A meta-analysis of the association between patients' early perception of treatment credibility and their posttreatment outcomes. *Psychotherapy*, *55*(4), 486–495. https://doi.org/10.1037/pst0000168

Constantino, M. J., Coyne, A. E., & Muir, H. J. (2020). Evidence-based therapist responsivity to disruptive clinical process. *Cognitive and Behavioral Practice*, *27*(4), 405–416. https://doi.org/10.1016/j.cbpra.2020.01.003

Conte, H. R. (1994). Review of research in supportive psychotherapy: An update. *American Journal of Psychotherapy*, *48*(4), 494–504. https://doi.org/10.1176/appi.psychotherapy.1994.48.4.494

Creed, T. A., Wolk, C. B., Feinberg, B., Evans, A. C., & Beck, A. T. (2016). Beyond the label: Relationship between community therapists' self-report of a cognitive behavioral therapy orientation and observed skills. *Administration and Policy in Mental Health*, *43*(1), 36–43. https://doi.org/10.1007/s10488-014-0618-5

Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: A meta-analysis. *Clinical Psychology Review*, *32*(4), 280–291. https://doi.org/10.1016/j.cpr.2012.01.003

Dinakar, K., Chen, J., Lieberman, H., Picard, R., & Filbin, R. (2015). Mixed-initiative real-time topic modeling & visualization for crisis counseling. *Proceedings of the 20th International Conference on Intelligent User Interfaces-IUI '15*, 417–426. https://doi.org/10.1145/2678025.2701395

Dunn, C., Darnell, D., Atkins, D. C., Hallgren, K. A., Imel, Z. E., Bumgardner, K., Owens, M., & Roy-Byrne, P. (2016). Within-provider variability in motivational interviewing integrity for three years after MI training: Does time heal? *Journal of Substance Abuse Treatment*, *65*, 74–82. https://doi.org/10.1016/j.jsat.2016.02.008

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, *48*(1), 43–49. https://doi.org/10.1037/a0022187

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy*, *55*(4), 399–410. https://doi.org/10.1037/pst0000175

Ewbank, M. P., Cummins, R., Tablan, V., Bateup, S., Catarino, A., Martin, A. J., & Blackwell, A. D. (2020). Quantifying the association between psychotherapy content and clinical outcomes using deep learning. *JAMA Psychiatry*, *77*(1), 35–43. https://doi.org/10.1001/jamapsychiatry.2019.2664

Flemotomos, N., Martinez, V. R., Chen, Z., Singla, K., Ardulov, V., Peri, R., Caperton, D. D., Gibson, J., Tanana, M. J., Georgiou, P., Van Epps, J., Lord, S. P., Hirsch, T., Imel, Z. E., Atkins, D. C., & Narayanan, S. (2022). Automated evaluation of psychotherapy skills using speech and language technologies. *Behavior Research Methods*, *54*, 690–711. https://doi.org/10.3758/s13428-021-01623-4

Gibson, J., Can, D., Xiao, B., Imel, Z. E., Atkins, D. C., Georgiou, P., & Narayanan, S. S. (2016). *A deep learning approach to modeling empathy in addiction counseling* [Conference session]. Proc. Interspeech 2016. https://doi.org/10.21437/Interspeech.2016-554

Goldberg, S. B., Tanana, M., Imel, Z. E., Atkins, D. C., Hill, C. E., & Anderson, T. (2021). Can a computer detect interpersonal skills? Using machine learning to scale up the Facilitative Interpersonal Skills task. *Psychotherapy Research*, *31*(3), 281–288. https://doi.org/10.1080/10503307.2020.1741047

Greenberg, L. S., Elliott, R., Watson, J. C., & Bohart, A. C. (2001). Empathy. *Psychotherapy*, *38*(4), 380–384. https://doi.org/10.1037/0033-3204.38.4.380

Hallgren, K. A., Dembe, A., Pace, B. T., Imel, Z. E., Lee, C. M., & Atkins, D. C. (2018). Variability in motivational interviewing adherence across sessions, providers, sites, and research contexts. *Journal of Substance Abuse Treatment*, *84*, 30–41. https://doi.org/10.1016/j.jsat.2017.10.011

Hilbert, K., Kunas, S. L., Lueken, U., Kathmann, N., Fydrich, T., & Fehm, L. (2020). Predicting cognitive behavioral therapy outcome in the outpatient sector based on clinical routine data: A machine learning approach. *Behaviour Research and Therapy*, *124*, Article 103530. https://doi.org/10.1016/j.brat.2019.103530

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*(1), 9–16. https://doi.org/10.1037/a0022186

Houck, J., Moyers, T., Miller, W., Glynn, L., & Hallgren, K. (2010). *Motivational interviewing skill code (MISC) 2.5*. https://casaa.unm.edu/download/misc25.pdf

Imel, Z. E., Baer, J. S., Martino, S., Ball, S. A., & Carroll, K. M. (2011). Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence*, *115*(3), 229–236. https://doi.org/10.1016/j.drugalcdep.2010.11.010

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Imel, Z. E., Baldwin, S. A., Baer, J. S., Hartzler, B., Dunn, C., Rosengren, D. B., & Atkins, D. C. (2014). Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients. *Journal of Consulting and Clinical Psychology*, *82*(3), 472–481. https://doi.org/10.1037/a0036158

Imel, Z. E., Caperton, D. D., Tanana, M., & Atkins, D. C. (2017). Technology-enhanced human interaction in psychotherapy. *Journal of Counseling Psychology*, *64*(4), 385–393. https://doi.org/10.1037/cou0000213

Imel, Z. E., Pace, B. T., Soma, C. S., Tanana, M., Hirsch, T., Gibson, J., Georgiou, P., Narayanan, S., & Atkins, D. C. (2019). Design feasibility of an automated, machine-learning based feedback system for motivational interviewing. *Psychotherapy*, *56*(2), 318–328. https://doi.org/10.1037/pst0000221

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy research: Scaling up the evaluation of patient–provider interactions. *Psychotherapy*, *52*(1), 19–30. https://doi.org/10.1037/a0036841

Irving, L. M., Snyder, C. R., Cheavens, J., Gravel, L., Hanke, J., Hilberg, P., & Nelson, N. (2004). The relationships between hope and outcomes at the pretreatment, beginning, and later phases of psychotherapy. *Journal of Psychotherapy Integration*, *14*(4), 419–443. https://doi.org/10.1037/1053-0479.14.4.419

Jung, J.-W., Heo, H.-S., Shim, H.-J., & Yu, H.-J. (2019). *Short utterance compensation in speaker verification via cosine-based teacher-student learning of speaker embeddings* [Conference session]. 2019 IEEE Automatic Speech Recognition and Understanding Workshop. https://doi.org/10.1109/ASRU46091.2019.9004029

Kanagasundaram, A., Vogt, R., Dean, D., Sridharan, S., & Mason, M. (2011). *I-vector based speaker recognition on short utterances* [Conference session]. Proc. Interspeech 2011. https://doi.org/10.21437/Interspeech.2011-58

Karlin, B. E., & Cross, G. (2014). From the laboratory to the therapy room: National dissemination and implementation of evidence-based psychotherapies in the U.S. Department of Veterans Affairs Health Care System. *American Psychologist*, *69*(1), 19–33. https://doi.org/10.1037/a0033888

Katz, M., Hilsenroth, M., Moore, M., & Gold, J. R. (2021). Profiles of adherence and flexibility in psychodynamic psychotherapy: A cluster analysis. *Journal of Psychotherapy Integration*, *31*(4), 348–362. https://doi.org/10.1037/int0000226

Katz, M., Hilsenroth, M. J., Gold, J. R., Moore, M., Pitman, S. R., Levy, S. R., & Owen, J. (2019). Adherence, flexibility, and outcome in psychodynamic treatment of depression. *Journal of Counseling Psychology*, *66*(1), 94–103. https://doi.org/10.1037/cou0000299

Koller, M. (2016). robustlmm: An R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software*, *75*(6), 1–24. https://doi.org/10.18637/jss.v075.i06

Li, X., Kivlighan, D. M., & Hill, C. E. (2020). Follow you or follow me? Examining therapist responsiveness to client and responsiveness to self, using differential equations model and multilevel data disaggregation from an interpersonal theory framework. *Journal of Counseling Psychology*, *67*(5), 608–621. https://doi.org/10.1037/cou0000406

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(59), Article 3112. https://doi.org/10.21105/joss.03139

Luo, Y., Ye, J., Adams, R. B., Jr., Li, J., Newman, M. G., & Wang, J. Z. (2020). ARBEE: Towards automated recognition of bodily expression of emotion in the wild. *International Journal of Computer Vision*, *128*(1), 1–25. https://doi.org/10.1007/s11263-019-01215-y

Magill, M., Apodaca, T. R., Borsari, B., Gaume, J., Hoadley, A., Gordon, R. E. F., Tonigan, J. S., & Moyers, T. (2018). A meta-analysis of motivational interviewing process: Technical, relational, and conditional process models of change. *Journal of Consulting and Clinical Psychology*, *86*(2), 140–157. https://doi.org/10.1037/ccp0000250

Magill, M., Gaume, J., Apodaca, T. R., Walthers, J., Mastroleo, N. R., Borsari, B., & Longabaugh, R. (2014). The technical hypothesis of motivational interviewing: A meta-analysis of MI's key causal model. *Journal of Consulting and Clinical Psychology*, *82*(6), 973–983. https://doi.org/10.1037/a0036833

Maskit, B. (2021). Overview of computer measures of the referential process. *Journal of Psycholinguistic Research*, *50*(1), 29–49. https://doi.org/10.1007/s10936-021-09761-8

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. *American Psychologist*, *65*(2), 73–84. https://doi.org/10.1037/a0018121

Nakagawa, S., & Schielzeth, H. (2017). *Extending R2 and intra-class correlation coefficient from generalized linear mixed-effects models: Capturing and characterizing biological variation*. BioRxiv. https://doi.org/10.1101/095851

Nienhuis, J. B., Owen, J., Valentine, J. C., Winkeljohn Black, S., Halford, T. C., Parazak, S. E., Budge, S., & Hilsenroth, M. (2018). Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: A meta-analytic review. *Psychotherapy Research*, *28*(4), 593–605. https://doi.org/10.1080/10503307.2016.1204023

Owen, J., & Hilsenroth, M. J. (2014). Treatment adherence: The importance of therapist flexibility in relation to therapy outcomes. *Journal of Counseling Psychology*, *61*(2), 280–288. https://doi.org/10.1037/a0035753

Pace, B. T., Dembe, A., Soma, C. S., Baldwin, S. A., Atkins, D. C., & Imel, Z. E. (2017). A multivariate meta-analysis of motivational interviewing process and outcome. *Psychology of Addictive Behaviors*, *31*(5), 524–533. https://doi.org/10.1037/adb0000280

Pace, B. T., Tanana, M., Xiao, B., Dembe, A., Soma, C. S., Steyvers, M., Narayanan, S., Atkins, D. C., & Imel, Z. E. (2016). What about the words? Natural language processing in psychotherapy. *Psychotherapy Bulletin*, *51*(1), 14–18. https://societyforpsychotherapy.org/wp-content/uploads/2014/11/Bulletin51-1_Online.pdf

Peabody, S. A., & Gelso, C. J. (1982). Countertransference and empathy: The complex relationship between two divergent concepts in counseling. *Journal of Counseling Psychology*, *29*(3), 240–245. https://doi.org/10.1037/0022-0167.29.3.240

Probst, G. H., Berger, T., & Flückiger, C. (2019). The alliance-outcome relation in internet-based interventions for psychological disorders: A correlational meta-analysis. *Verhaltenstherapie*. Advance online publication. https://doi.org/10.1159/000503432

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rockland, L. H. (1993). A review of supportive psychotherapy, 1986–1992. *Hospital & Community Psychiatry*, *44*(11), 1053–1060. https://doi.org/10.1176/ps.44.11.1053

Schwalbe, C. S., Oh, H. Y., & Zweben, A. (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction*, *109*(8), 1287–1294. https://doi.org/10.1111/add.12558

Snippe, E., Schroevers, M. J., Tovote, K. A., Sanderman, R., Emmelkamp, P. M. G., & Fleer, J. (2019). Explaining variability in therapist adherence and patient depressive symptom improvement: The role of therapist interpersonal skills and patient engagement. *Clinical Psychology & Psychotherapy*, *26*(1), 84–93. https://doi.org/10.1002/cpp.2332

Stiles, W. B., & Horvath, A. O. (2017). Appropriate responsiveness as a contribution to therapist effects. In L. G. Castonguay & C. E. Hill (Eds.), *How and why are some therapists better than others?: Understanding therapist effects* (pp. 71–84). American Psychological Association. https://doi.org/10.1037/0000034-005

Tanana, M. J., Soma, C. S., Kuo, P. B., Bertagnolli, N. M., Dembe, A., Pace, B. T., Srikumar, V., Atkins, D. C., & Imel, Z. E. (2021). How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behavior Research Methods*, *53*(5), 2069–2082. https://doi.org/10.3758/s13428-020-01531-z

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*(3), 457–469. https://doi.org/10.1177/2167702617691560

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). Routledge. https://doi.org/10.4324/9780203582015

Watson, J. C., & Wiseman, H. E. (2021). *The responsive psychotherapist: Attuning to clients in the moment*. American Psychological Association. https://doi.org/10.1037/0000240-000

Wiltsey Stirman, S., Calloway, A., Toder, K., Miller, C. J., Devito, A. K., Meisel, S. N., Xhezo, R., Evans, A. C., Beck, A. T., & Crits-Christoph, P. (2013). Community mental health provider modifications to cognitive therapy: Implications for sustainability. *Psychiatric Services*, *64*(10), 1056–1059. https://doi.org/10.1176/appi.ps.201200456

Wiltsey Stirman, S., Gutner, C. A., Crits-Christoph, P., Edmunds, J., Evans, A. C., & Beidas, R. S. (2015). Relationships between clinician-level attributes and fidelity-consistent and fidelity-inconsistent modifications to an evidence-based psychotherapy. *Implementation Science : IS*, *10*(1), Article 115. https://doi.org/10.1186/s13012-015-0308-z

Winston, A., Pinsker, H., & McCullough, L. (1986). A review of supportive psychotherapy. *Hospital & Community Psychiatry*, *37*(11), 1105–1114. https://doi.org/10.1176/ps.37.11.1105

Wiser, S., & Goldfried, M. (1996). Verbal interventions in significant psychodynamic-interpersonal and cognitive-behavioral therapy sessions. *Psychotherapy Research*, *6*(4), 309–319. https://doi.org/10.1080/10503309612331331828

Xiao, B., Can, D., Georgiou, P. G., Atkins, D., & Narayanan, S. S. (2012). *Analyzing the language of therapist empathy in motivational interview based psychotherapy* [Conference session]. Proceedings of The 2012 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference.

Xiao, B., Huang, C., Imel, Z. E., Atkins, D. C., Georgiou, P., & Narayanan, S. S. (2016). A technology prototype system for rating therapist empathy from audio recordings in addiction counseling. *PeerJ Computer Science*, *2*, Article e59. https://doi.org/10.7717/peerj-cs.59

Xiao, B., Imel, Z. E., Atkins, D. C., Georgiou, P. G., & Narayanan, S. S. (2015). *Analyzing speech rate entrainment and its relation to therapist empathy in drug addiction counseling* [Conference session]. Proc. Interspeech 2015. https://doi.org/10.21437/Interspeech.2015-537

Xiao, B., Imel, Z. E., Georgiou, P. G., Atkins, D. C., & Narayanan, S. S. (2015). "Rate my therapist": Automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLOS ONE*, *10*(12), Article e0143055. https://doi.org/10.1371/journal.pone.0143055

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, *19*(4), 399–404. https://doi.org/10.1111/j.1467-9280.2008.02099.x

Received April 3, 2022
Revision received August 22, 2022
Accepted August 23, 2022 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!