Psychotherapy

Automating the Assessment of Multicultural Orientation Through Machine Learning and Natural Language Processing

Simon B. Goldberg, Michael Tanana, Shaakira Haywood Stewart, Camille Y. Williams, Christina S. Soma, David C. Atkins, Zac E. Imel, and Jesse Owen

Online First Publication, February 1, 2024. https://dx.doi.org/10.1037/pst0000519

Psychotherapy

ISSN: 0033-3204
https://doi.org/10.1037/pst0000519

Automating the Assessment of Multicultural Orientation Through Machine Learning and Natural Language Processing

Simon B. Goldberg[1, 2], Michael Tanana[3], Shaakira Haywood Stewart[4], Camille Y. Williams[1, 2], Christina S. Soma[3], David C. Atkins[3], Zac E. Imel[3, 5], and Jesse Owen[4]

[1] Department of Counseling Psychology, University of Wisconsin-Madison
[2] Center for Healthy Minds, University of Wisconsin-Madison
[3] Lyssn.io, Seattle, Washington, United States
[4] Department of Counseling Psychology, University of Denver
[5] Department of Educational Psychology, University of Utah

Recent scholarship has highlighted the value of therapists adopting a multicultural orientation (MCO) within psychotherapy. A newly developed performance-based measure of MCO capacities exists (MCO–performance task [MCO-PT]) in which therapists respond to video-based vignettes of clients sharing culturally relevant information in therapy. The MCO-PT provides scores related to the three aspects of MCO: cultural humility (i.e., adoption of a nonsuperior and other-oriented stance toward clients), cultural opportunities (i.e., seizing or making moments in session to ask about clients' cultural identities), and cultural comfort (i.e., therapists' comfort in cultural conversations). Although a promising measure, the MCO-PT relies on labor-intensive human coding. The present study evaluated the ability to automate the scoring of the MCO-PT transcripts using modern machine learning and natural language processing methods. We included a sample of 100 participants ($n = 613$ MCO-PT responses). Results indicated that machine learning models were able to achieve near-human reliability on the average across all domains (Spearman's $\rho = .75$, $p < .0001$) and opportunity ($\rho = .81$, $p < .0001$). Performance was less robust for cultural humility ($\rho = .46$, $p < .001$) and was poorest for cultural comfort ($\rho = .41$, $p < .001$). This suggests that we may be on the cusp of being able to develop machine learning-based training paradigms that could allow therapists opportunities for feedback and deliberate practice of some key therapist behaviors, including aspects of MCO.

Clinical Impact Statement

Question: Can a video-based performance task assessing multicultural orientation be automatically scored using machine learning? ***Findings:*** Machine learning models achieved near-human reliability in some domains, although performance was more modest in other domains. ***Meaning:*** As machine learning methods continue to advance and larger data sets are available for model training, machine learning methods may be able to automate the scoring of tasks previously dependent on human coders. ***Next Steps:*** We may be on the cusp of being able to develop automated training paradigms that provide specific behavioral feedback to aid in the development of multicultural orientation.

Keywords: multicultural orientation, multicultural competence, machine learning, natural language processing, technology

Supplemental materials: https://doi.org/10.1037/pst0000519.supp

Editor's Note. James F. Boswell served as the guest editor for this article.—JO

Simon B. Goldberg https://orcid.org/0000-0002-6888-0126

Michael Tanana, David C. Atkins, and Zac E. Imel are cofounders with equity stake in a technology company, Lyssn.io, focused on tools to support training, supervision, and quality assurance of psychotherapy and counseling. Jesse Owen has vested in Lyssn.io. The remaining authors report no conflicts of interest. This research was supported by the National Center for Complementary and Integrative Health Grant K23AT010879 awarded to Simon B. Goldberg, the National Institute of Mental Health of the National Institutes of Health under Grant T32MH018931-31 awarded to Camille Y. Williams and Grant R42MH128101 awarded to David C. Atkins, and the John Templeton Foundation Grant 61603 awarded to Jesse Owen. The content reported in this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

There is a currently in press work stemming from this data set (Stewart et al., 2023).

Simon B. Goldberg played a lead role in writing–original draft, a supporting role in data curation, investigation, and methodology, and an equal role in conceptualization and funding acquisition. Michael Tanana played a lead role in formal analysis and software, a supporting role in writing–original draft, and an equal role in data curation, methodology, validation, and writing–review and editing. Shaakira Haywood Stewart played a supporting role in conceptualization, data curation, project administration, writing–original draft, and writing–review and editing and an equal role in investigation, methodology, and resources. Camille Y. Williams played a supporting role in

continued

The importance of multicultural competencies within psychotherapy has been recognized for over a half of a century (e.g., Pine, 1972), with renewed recognition in the era of Black Lives Matter and contemporary antiracism movements (e.g., Hargons et al., 2017; Kendi, 2019; Roberts & Rizzo, 2021). Multicultural competencies include a therapist's ability to effectively apply multicultural awareness, knowledge, and skills within therapy (American Psychological Association, 2003). More recent scholarship on multiculturalism in psychotherapy has emphasized the notion of adopting a multicultural orientation (MCO; J. Owen, 2013; J. J. Owen, Tao, et al., 2011; Hook et al., 2016). Complementary to multicultural competency, MCO is characterized as a way of being in therapy (J. Owen, Leach, et al., 2011; J. J. Owen, Tao, et al., 2011) and includes three components: (a) cultural humility (i.e., adoption of a nonsuperior and other-oriented stance toward clients; J. Owen, 2013; Hook et al., 2016), (b) cultural opportunities (i.e., seizing or making moments in session to ask about clients' cultural identities), and (c) cultural comfort (i.e., therapists' comfort in cultural conversations; Pérez-Rojas et al., 2019).

It is increasingly clear that therapists' ability to work with cultural factors in psychotherapy is important. Measures of therapists' multicultural competence have shown moderate magnitude associations with treatment outcomes ($r = .29$) and large magnitude associations with key treatment process variables such as therapeutic alliance ($r = .61$) and session depth ($r = .58$; Tao et al., 2015; see also Soto et al., 2018). Similarly, MCO constructs have been associated with stronger alliances and therapy outcomes in over 20 studies with nearly 10,000 clients (for reviews, see Davis et al., 2018; Zhang et al., 2022). In contrast, microaggressions or "subtle, stunning, often automatic, and non-verbal exchanges which are 'put downs'" (Pierce et al., 1977, p. 66) have been documented in therapy (e.g., Hook et al., 2016; Owen, Imel, et al., 2011). Indeed, microaggressions are common for racial/ethnic minority clients and contribute to broader health care disparities across racial/ethnic groups (Ehie et al., 2021; J. Owen et al., 2019). Importantly, clients who rate their therapist higher on cultural humility, one of the MCO pillars, report fewer microaggressions (Davis et al., 2016; DeBlaere et al., 2023). Thus, it is imperative to develop methods to equip therapists to effectively work with cultural elements in therapy.

To effectively train therapists to adopt MCO within therapy, it is important to develop methods to assess MCO accurately. As with the development of expertise in various professional domains, including psychotherapy, the availability of accurate feedback on one's performance is essential for improvement (Kahneman & Klein, 2009; Tracey et al., 2014). Such feedback can allow therapists to modify their behavior to improve their performance. To date, the bulk of the research examining multicultural competence and MCO within psychotherapy has relied on client or therapist report (e.g., multicultural awareness/knowledge/skills survey, Cultural Humility Scale; Cross-Cultural Competencies Inventory–Revised; D'Andrea et al., 1991; Drinane et al., 2016; Hook et al., 2013; see also Tao et al., 2015). As discussed by Tracey et al. (2014), there are a host of limitations associated with relying primarily on clients' or therapists' self-report that may make such assessments misleadingly affirmative. Thus, there is value in developing more objective methods for assessing multicultural competence and MCO within psychotherapy.

Progress has been made in developing objective assessments of key therapist skills. Perhaps the most developed assessment of this kind is the facilitative interpersonal skills (FIS) task (Anderson et al., 2009). The FIS is a performance-based task in which therapists are provided with a series of vignettes depicting interpersonally challenging therapy interactions. The therapists are instructed to respond to the vignettes as if they were the client's therapist. Therapists' responses are then evaluated by trained coders across eight domains related to interpersonal skills (e.g., verbal fluency, persuasiveness, emotional expression, and empathy; Anderson et al., 2009). The FIS has emerged as one of the few measures of therapist characteristics that have been consistently shown to predict client outcomes (e.g., Anderson et al., 2009, 2016; Heinonen & Nissen-Lie, 2020). In theory, performance-based tasks like the FIS can provide assessment of therapist skills that are less vulnerable to know biases in self-report measures commonly used in psychotherapy (e.g., social desirability, ceiling effects; Goldberg et al., 2023; Tracey, 2016).

Recent progress has also been made in the development of a performance-based task designed to assess therapist MCO. Like the FIS,[1] the MCO performance task (MCO-PT; Stewart et al., 2023) provides therapists with a series of eight video-based vignettes depicting psychotherapy interactions (note only seven videos are coded, the first video is a practice video). Instead of capturing interpersonally challenging moments, the MCO-PT vignettes include instances of clients sharing multiculturally relevant content with the therapist. Therapists' responses are then evaluated by trained coders across four domains. These domains include (a) cultural humility (i.e., maintenance of an "other-oriented" stance that is characterized by nonsuperiority and curiosity), cultural comfort (i.e., ability to engage with cultural content with ease, confidence, and openness), cultural opportunities (i.e., awareness of and attunement to cultural cues, taking opportunities to discuss cultural content), and (b) an overall rating reflecting the degree to which therapists' responses were aimed to help the client feel validated and connected to the therapist (i.e., a good response) versus responses that were not connected to the client, were insulting, or were otherwise inappropriate (i.e., a bad response). To date, the MCO-PT has shown promising psychometric properties,

[1] Note that the FIS has been recently applied to evaluate facilitative interpersonal skills when responding to cultural content as well (see Schwartzman, 2022).

conceptualization, funding acquisition, and writing–review and editing and an equal role in writing–original draft. Christina S. Soma played a supporting role in conceptualization, project administration, and writing–review and editing. David C. Atkins played a supporting role in conceptualization, formal analysis, funding acquisition, project administration, resources, software, supervision, and writing–review and editing. Zac E. Imel played a supporting role in investigation, methodology, project administration, software, supervision, and writing–review and editing and an equal role in conceptualization, formal analysis, and funding acquisition. Jesse Owen played a supporting role in data curation, formal analysis, investigation, methodology, supervision, writing–original draft, and writing–review and editing and an equal role in conceptualization, funding acquisition, and project administration.

Correspondence concerning this article should be addressed to Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin-Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI 53706, United States. Email: sbgoldberg@wisc.edu

including consistently high intraclass correlations (ICCs), three-factor solution representing the three MCO pillars, and high Cronbach's αs (Stewart et al., 2023).

Were it implemented at scale, performance tasks like the MCO-PT could provide a valuable objective assessment that could be used to guide therapists' development of MCO within psychotherapy. However, implementation of the MCO-PT and other performance-based measures of therapist capacities is currently limited. Training coders to reliability takes approximately 8 hr per coder for the MCO-PT and coding responses is also highly time-consuming (e.g., 3–5 min per response for seven videos or 21–35 min per participant). There have been substantial barriers to the implementation of observer-rated coding of responses to vignettes and therapy sessions for decades in psychotherapy research (Atkins et al., 2014; Tanana et al., 2016).

Natural language processing (NLP) is a broad subfield of machine learning (ML) in which statistical programs can learn patterns from unstructured text in ways that allow computer programs to interpret or generate human language (Jurafsky & Martin, 2023, see Aafjes-van Doorn et al., 2021, for discussion of these methods specifically in the context of psychotherapy). Recent advances in NLP may greatly expand access to these tools in psychotherapy training and research by obviating the need for resource-intensive human coding. Currently, ML and NLP have shown promise for automating several other psychotherapy process variables. For example, a series of studies have demonstrated that adherence to motivational interviewing can be objectively assessed through analysis of therapist speech, with ML models showing reliability similar to human coders (e.g., Atkins et al., 2014; Flemotomos et al., 2022; Tanana et al., 2016; Xiao et al., 2015). Other work has detected client-rated therapeutic alliance from psychotherapy session recordings (Goldberg, Flemotomos, et al., 2020) and linked topics discussed in psychotherapy with changes in client distress (Atzil-Slonim et al., 2021).

ML and NLP have also been used to automate the scoring of the FIS (Goldberg et al., 2021). In a sample of 164 undergraduates who completed the FIS task, ML models predicted FIS scores above chances (ρs = .27–.53) and achieved 31%–60% of human reliability. For example, human reliability on the FIS total score was ICC = .93 and the ML models achieved ρ = .48, which is 52% of the human reliability (i.e., .48/.93). Comparison with human rater reliability is important as human raters in theory provide an upper bound for reliability that ML models can achieve (Atkins et al., 2014). Although a promising first attempt at automating the scoring of psychotherapy performance tasks, this level of reliability is still below cutoffs typically recommended for use in clinical and research settings (i.e., $r = .70$; Lance et al., 2006). However, recent advances in ML and NLP methods have shown promise for improving model performance. One particularly promising approach for extracting text for analyses is the Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach (RoBERTa; Liu et al., 2019). This algorithm, which is discussed in greater detail in the Method section, breaks down sentences into words and word pairs (i.e., tokenization), thereby striking a balance between word-based language models and character-based language models (e.g., Sennrich et al., 2015). In theory, this approach can provide a more robust application to psychotherapy-specific language structures and vocabulary. RoBERTa is able to move beyond simpler "bag of words" approaches that examine small groupings of words (i.e., *n*-gram models, focusing on single words [unigram], two words [bigram], etc.) and instead allows words to be evaluated within their context. Preliminary evidence suggests that RoBERTa may dramatically improve the accuracy of ML models applied to psychotherapy processes (e.g., Flemotomos et al., 2021).

The Present Study

The present study sought to develop automated ML-based scoring of the MCO-PT. The development of such a method that is not dependent on human coders would make possible the scaling up of MCO-PT, a construct with high clinical relevance particularly when working with clients holding minoritized identities (J. J. Owen, Tao, et al., 2011). In addition to research applications, accurate automated assessment of MCO would support the possibility of eventually building training paradigms that provide automated feedback related to MCO. To explore the possibility of automating MCO-PT, we used a sample of MCO-PT assessments that were scored by a team of trained human coders. We then examined the performance of ML models using RoBERTa. These analyses were exploratory and the study was not preregistered.

Method

Participants

Data were drawn from a previously conducted study validating the MCO-PT (Stewart et al., 2023). A total of 74 graduate students enrolled in counseling and other professional psychology and related professional training programs participated in the study. A total of 26 undergraduate students were recruited to maximize the variance in responses. Approximately 80% of the participants self-identified as non-Latinx White, with the remaining participants identifying as Black (4%), Latinx (6%), Eastern Asian (4%), Middle Eastern (1%), and biracial/multiracial (5%). Participant ages ranged from 18 to 40 years old ($M = 24.35$, $SD = 4.99$). Most participants self-identified as cisgender female (82%), with the remaining participants identifying as cisgender male (13%), transgender/gender nonconforming (1%), or not reporting their gender identity (4%). Most participants self-identified as heterosexual/straight (81%), with the remaining participants identifying as lesbian, gay, bisexual, transgender, queer, others (15%), or not reporting their sexual orientation (4%).

There was a total of nine coders for the present study, eight of whom identified as cisgender female, and one cisgender male. Seven of the nine coders self-identified as non-Latinx White, one identified as South Asian, and one identified as African American. The team of coders were trained for approximately 8 hr. They were provided with information about the MCO framework, scholarly articles on MCO, and instruction on how to use the coding form. Coders coded practice videos until they consistently reached an ICC coefficient of .70 (see Stewart et al., 2023, for further details). The coding training took approximately 15 hr per coder.

MCO Performance Task (Stewart et al., 2023)

This performance-based task was designed to elicit responses to client videos with content that included aspects of clients' cultural identities. To create the vignettes, culturally diverse graduate students

and a professor in a counseling psychology drafted content based on their personal and professional experiences in psychotherapy. The scripts varied in sociocultural diversity (e.g., race/ethnicity, gender identity, sexual orientation, religion, ability status), presenting concerns (e.g., depression, anxiety, substance abuse, relationship difficulties, identity development, adjustment/transition), phase in therapy (though most were early), and degree of clinical difficulty. Lower clinical difficulty clips ended with a client statement or with a general question back to the therapist (e.g., "What do you think?"). Higher clinical difficulty clips elicited value statements from the therapist (e.g., "Is it me or is it all gay men?").

There was a total of eight videos (one video was for practice and seven were used for analysis). All participants received the same practice video first, and the subsequent videos were randomized to account for order effects. The final MCO-PT included the following seven clients talking for ~1–2 min each: (a) Anthony, a 26-year-old, Black cis man experiencing discrimination and isolation as the only Black man in his office, (b) Jasmine, a 19-year-old White cis woman experiencing social anxiety and judgment about the way her Indian boyfriend's family shares meals at dinner, (c) Julie, a 24-year-old Chinese American trans woman having relationship difficulties, (d) Aleemah, a 25-year-old Middle Eastern/American Muslim cis woman experiencing depression who states that the other doctoral students in her chemistry program do not understand her, (e) Stephen, a 43-year-old, White, Catholic, gay, cis man struggling with dating and maintaining his sobriety, (f) Arlene, a 45-year-old White cis woman experiencing postpartum depression after the recent birth of her daughter and judgment from others about her age, and (g) Cathy, a 25-year-old White, gender fluid person with anxiety and depression related to workplace discrimination. The actors who portrayed these characters were either graduate students or university faculty volunteers.

As described in Stewart et al. (2023), the MCO-PT is scored across three primary domains with raters assessing items on a 6-point scale. Scale anchors reflect low to high levels of the target construct. Anchors from the four domains include: cultural comfort (low = uncomfortable, high = comfortable), cultural humility (low = disrespectful, high = respectful), cultural opportunity (low = no cultural discussion, high = definitive cultural discussion). Internal consistency reliability estimates for the MCO-PT were .94 for cultural comfort, .97 for cultural humility, and .77 for cultural opportunities. The MCO-PT also demonstrated a three-factor structure mirroring the MCO core pillars.

Our approach was modeled after Anderson et al. (2009) in that participants were asked to respond in real time to clients who were filmed facing the camera while describing their culturally linked presenting concerns. Participants were directed to respond to the client vignettes using their webcam as if they were talking directly to the client actor. They were only given one chance to respond to each video and responses were recorded via the online platform Skillsetter, Inc (https://www.skillsetter.com).

Procedure

This study was approved by the Institutional Review Board at University of Denver. Recruitment messages were sent to a clinical mental health master's program listserv in the Mountain West. The announcement included a brief description of the study and a link to participate. After completing the MCO-PT, participants were directed to a Qualtrics survey, where demographic and other self-report data were collected. Participants were offered extra credit for their courses to participate. Undergraduate participants were offered credit for completing the study via the research requirements for their degree programs.

Data Analysis

Following Kuo et al. (2023), we prepared human transcribed text for analysis using the "RoBERTa" byte-pair encoding (Liu et al., 2019). RoBERTa is a transformer neutral network that converts sentences or paragraphs into numeric vectors that can be used for prediction tasks. Transformer models are a class of ML approaches that rely on pretrained language models that can then be "fine-tuned" on specific labeled data. Specifically, as a more robust version of its predecessor, Bidirectional Encoder Representations from Transformers (Devlin et al., 2018), RoBERTa is pretrained on a large corpus (160 GB) of English text to predict words that are hidden from the model. The base-English pretrained model was used in our analyses, which includes 355 million parameters (hidden size of 1,024, 16 attention heads, and 24 layers, see Bishop, 2006, for explanation of these aspects of transformer models). We used the base model rather than the large model to limit video random-access memory burden used for each training example. Video random-access memory is a form of computer memory which impacts the processing time necessary for computationally intensive models. Transcripts of participant responses to the MCO-PT were used for analysis. As noted, RoBERTa breaks sentences into words and word pairs, using a 50,000-word vocabulary. Inputs were wrapped in the standard RoBERTa starting and ending tags ("<s>" and "</s>"). For the current analyses, the pretrained model was downloaded from the "Huggingface" repository (Wolf et al., 2019).

Models were trained using the adaptive moment estimation (Adam) optimizer which guides how the models "learn" during training (see Kingma & Ba, 2014). The models were customized to aggregate the classifier token (i.e., word or word-pair outputs) into a convolutional neural network (i.e., an unsupervised pattern classification neural network) that feeds into a linear predictor. In addition to this hierarchical model enabling prediction of longer sequences, the linear predictor also enabled us to predict continuous variables, and thus utilize a mean square error loss model to measure model accuracy (i.e., the discrepancy between a word or word-pair's true value and the model's predicted value). Data were divided into 80% for training, 10% for the validation set used to tune the learning rate, and 10% for a held-out test set to evaluate performance of the best fitting model. To avoid artificially inflating model performance, we ensured that participants' responses only appeared in one of the three divisions (i.e., participants with responses in the training set did not appear in the validation or held-out test set). We used the "Pytorch" (v. 1.1.1; Paszke et al., 2017) and "Huggingface" (Wolf et al., 2019) frameworks in Python to train the models. Models were trained on an Nvidia Quadro 8,000 with 48 GB of video random-access memory. We used the validation set to tune the number of epochs. Spearman's ρ (with a value of 1 being perfect performance) was used as our main measure of model performance. We used R^2 as a measure of absolute performance within the held-out test set. Of note, this R^2 can theoretically be negative, since parameter estimates were derived outside of the test set. Models were constructed

predicting the three primary MCO domains (cultural comfort, cultural humility, cultural opportunity) as well as the average across these three domains.

Results

Descriptive Statistics

There was a potential of 700 responses (i.e., responses to seven vignettes from $n = 100$ participants); however, 87 were either skipped by participants or were recordings of insufficient quality to allow transcription. All participants had scores on multiple vignettes (e.g., no participant skipped all the videos) and the average number of videos per participant was 6.13 (of seven). The correlation, derived from human coders, between cultural humility and cultural comfort was $r = .48$, $p < .001$, the correlation between cultural humility and cultural opportunities was $r = .43$, $p < .001$, and the correlation between cultural comfort and cultural opportunities was $r = .23$, $p < .05$. Human coders achieved acceptable reliability within each subscale (Table 1). The coding teams' average ICC across all domains was .85 (Table 1). The lowest ICC across all videos and coding teams was .54 (see Supplemental Table 1).

Machine Learning

Results from the RoBERTa models are displayed in Table 1. Mean square errors ranged from 0.39 (average across domains) to 1.06 (comfort). R^2 values ranged from .73 (cultural opportunity) to .15 (cultural comfort). In all cases, the RoBERTa models performed well above chance ($p < .001$). Model performance was best for the average across domains ($\rho = .75$, $p < .0001$) and cultural opportunity ($\rho = .81$, $p < .0001$). Model performance on cultural humility ($\rho = .46$, $p < .001$) and cultural comfort ($\rho = .41$, $p < .001$) was more modest. The RoBERTa models achieved a large proportion of the human coders' reliability (i.e., ICCs) for cultural opportunity (98.8%) and the average across domains (88.2%). The RoBERTa models achieved a smaller proportion of the human coders' reliability for cultural comfort (48.8%) and cultural humility (51.7%).

Discussion

The present study employed modern ML and NLP methods to automate coding of a performance-based task to assess therapist adoption of MCO—the MCO-PT (Stewart et al., 2023). This study adds to the growing literature showing the promise of ML and NLP for developing tools to assess aspects of psychotherapy process and outcome (Aafjes-van Doorn et al., 2021). Results from the RoBERTa-based models were particularly promising at least for some domains. These models demonstrated reliability similar to that of human coders for the cultural opportunity domain and the average across all three domains for which the RoBERTa models achieved >88% of human reliability. Model performance for the cultural comfort and cultural humility domains was more modest (48.8% and 51.7% of human coders' reliability). Of note, human coder-based reliability likely sets an upper limit on performance that can be achieved using ML (Atkins et al., 2014). Higher performance in the cultural opportunity domain may be due to the presence of clear language markers of culturally relevant content (e.g., references to aspects of the client's identity) whereas assessment of cultural comfort and cultural humility may involve consideration of nonverbal behavior and nonlinguistic speech features that were not available to the ML model. Nonetheless, model performance can be expected to improve as the available training corpus grows (Dwyer et al., 2018), which may eventually produce acceptable reliability even for cultural comfort and cultural humility. The inclusion of nonverbal and nonlinguistic speech features (e.g., prosody) may further improve model performance. Thus, it is conceivable that automated scoring of the MCO-PT may be ready for use in the coming years.

There are several key ways in which an automated MCO-PT could be used in practice. Within clinical settings, the MCO-PT could be implemented as a quality assurance or quality improvement measure. Practicing clinicians could evaluate their adoption of MCO using the task and, when relevant based on their performance, receive training to augment their areas of lower performance. Graduate training programs could also use the MCO-PT within courses focused on multicultural topics. Ultimately, it would be ideal to integrate an automatically scored MCO-PT within an ML-based training paradigm where therapists interact with ML-based software that provides automated feedback on their responses. Such a paradigm could provide therapists with opportunities to engage in deliberate practice focused on developing MCO. Deliberate practice, defined as "individualized training activities … to improve specific aspects of an individual's performance through repetition and successive refinement" (Ericsson & Lehmann, 1996, pp. 278–279) has been proposed as a means for improving therapists' performance in psychotherapy (Rousmaniere, 2019; Rousmaniere et al., 2017). Opportunities to engage in deliberate practice of relevant therapy behaviors is important, given psychotherapy does

Table 1
RoBERTa Model Results

MCO domain	Mean absolute error	Mean square error	R^2	Spearman ρ	p value	Human coder ICC	% Human coder
Comfort	0.86	1.06	.15	.41	.0006	.84	48.8%
Humility	0.61	0.54	.39	.46	.0001	.89	51.7%
Opportunity	0.67	0.80	.73	.81	<.0001	.82	98.8%
All domains	0.52	0.39	.48	.75	<.0001	.85	88.2%

Note. All domains = average across all three MCO domains; MCO = multicultural orientation; human coder ICC = intraclass correlation coefficient from human coding teams (see Supplemental Table 1); % human coder = proportion (i.e., percentage) of human coder ICC achieved by RoBERTa model; RoBERTa = Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.

not necessarily provide key ingredients for skill acquisition (i.e., predictable environment, opportunities to learn from past behavior; Kahneman & Klein, 2009; Tracey et al., 2014) and psychotherapists do not appear to improve simply by gaining experience (Goldberg et al., 2016). A tool for practicing MCO without clients may be a particularly welcome method for engaging in MCO-specific deliberate practice, given the risk of perpetrating microaggressions during the course of developing MCO (Freetly Porter et al., 2023). The provision of specific behavioral feedback may allow for learning opportunities that are not typically available in psychotherapy generally (Tracey et al., 2014).

The appearance of Chat Generative Pretrained Transformer (ChatGPT; Open AI, 2022) in November 2022 has brought the potential of ML, NLP, and artificial intelligence generally into the public view. ChatGPT is a specific transformer-based large language model that is trained using very large amounts of text data. This training allows ChatGPT the capacity to generate text similar to what a human might produce (Kasneci et al., 2023). The potential applications of ChatGPT and similar technologies are hard to overstate. They are likely to transform both health care and education, impacting relatively simple tasks such as generating discharge summaries (Patel & Lam, 2023) as well as more complex tasks such as supporting language learning and professional training (Kasneci et al., 2023). Tools based on large language models may provide the fluency necessary for building ML-powered tools that can provide responsive, self-guided training opportunities for developing MCO. Of course, there are various ethical and user experience-related barriers that must be addressed first, such as considering the potential implications of providing inaccurate ML-based feedback (knowing that the models will never be 100% accurate) and determining how to best provide feedback to users that will be useful to them and will promote learning rather than discouragement, frustration, and confusion (Kasneci et al., 2023; Patel & Lam, 2023). These issues notwithstanding, it is possible that these tools can be successfully integrated into clinical training for MCO and various other therapy-relevant skills in the coming years.

With the advent of new technologies, it is appropriate to have some measure of healthy skepticism regarding their potential (Coppersmith et al., 2022). The use of the models developed in the present study as well as future training extensions using large language models are by definition limited by the training data upon which they are based. Like any other program of research, generalizability and replication will be crucial for earning the trust of (and eventually the implementation by) therapists and educators. Moreover, while these tools have the potential to enhance therapists' training in MCO, they should not be considered a replacement of human educators and clinical supervisors. Indeed, ML MCO feedback should not be used for evaluative purposes but rather a way to get some immediate feedback, which may or may not capture fully the therapist's response. Nonetheless, they may ultimately prove to be valuable (if imperfect) tools for expanding opportunities for feedback and deliberate practice within the course of developing one's MCO.

Limitations

The present study has several important limitations. First, we analyzed transcripts rather than audio or video recordings which may have increased model performance above what would be expected if directly analyzing recordings. Moving to direct assessment of audio or video files may decrease performance as errors are introduced during the automated transcription process. In this study, the transcripts were double checked for accuracy before the ML models were conducted. Ultimately, for real-world implementation, it will be important to assess performance without labor-intensive human transcription of MCO-PT responses. A second limitation related to our use of transcripts is that our models did not include nonspeech elements which may improve performance. As noted above, nonverbal behaviors (e.g., nodding, posture) and nonlinguistic speech features (e.g., prosody) may contain important signals that were available to the human coders but not available to the ML model. Third, our sample size was modest, particularly by ML standards. Although perhaps acceptable for a proof-of-concept evaluation of automating MCO-PT scoring, future studies will ideally use a larger sample of coded responses. As noted, simply increasing the training set will likely yield improved performance in the test set (Dwyer et al., 2018). Additionally, our sample of respondents was majority non-Latinx White participants, which may limit generalizability. Having more responses from racially/ethnically diverse participants will be essential moving forward with this technology (Kostick-Quenet et al., 2022). Fourth, our sample included relatively few low-performing examples, which may have reduced the model's ability to reliably detect low-scoring responses. Future studies will ideally include a wider range of abilities, including low-scoring responses (e.g., overtly offensive comments). Fifth, although we conducted internal validation (i.e., cross-validation in a held-out test set), we did not conduct external validation (i.e., test of model performance in a completely independent data set). This will be important to do in the future to assess how the model will perform in data collected in different contexts.

Future Research Directions

There are many future directions to build on this early but promising work on automating the MCO-PT. First, it will be important to collect more data from more diverse samples to ensure that the NLP/ML models are robust, as any ML tool is only as good as the data it is trained on. Second, another viable area of research would be to connect the ML codes to therapy processes (e.g., alliance) and therapy outcomes. This could be done in many ways, such as examining how automated MCO-PT scores relate to therapist disparities in therapy outcomes (i.e., therapist effects). Third, there are many scenarios not captured within these videos. Thus, expanding the library of videos and developing coded responses for additional videos could be useful for both training and research.

Beyond ongoing model development and validation work, there are two critical potential applications of the ML models described here. As noted above, the first research application might use ML feedback as a training tool. There is an increasing focus on training for therapists that uses feedback that is proximal to practice (e.g., through use of the FIS and similar performance-based tasks, through evaluation of therapists' fidelity to a treatment specific; Allen et al., 2023; Goldberg, Baldwin, et al., 2020). At present, this work typically relies on the skill and reliability of human trainers. A key question will be first to do determine if training methods where therapists receive ML-based feedback are comparable in terms of

outcomes (both skill development and impact on client symptoms) to human provided feedback. Subsequent research might investigate issues related to implementation and cost effectiveness of ML-based training methods. For example, even if a training where a human provides feedback is 10% more efficacious than ML-based training, the impact on overall skill development of the mental health workforce for ML-based training might be orders of magnitude larger as the ML-based training's reach is not limited by the need to involve human coders and human trainers.

Second, an application not noted above is the use of ML models as screeners for therapist skill during selection for graduate school, internship, or professional staff positions. At present, these selection processes rely largely on resumes, recommendations, written communication (e.g., cover letters), and interviews, methods which have questionable predictive validity relative to skillful practice of psychotherapy (e.g., Schöttke et al., 2017). Future research could utilize scalable ML tools that evaluate therapist skill and examine if performance on these tasks predicts client outcomes like satisfaction, retention, and dropout—thus providing a more valid means of selecting therapists for professional roles.

Conclusion

There is increasing evidence that ML and NLP methods may be valuable for scaling up innovative methods within clinical practice and psychotherapy research. Results from the present study suggest that we may be on the cusp of being able to develop scalable training paradigms that could integrate deliberate practice into psychotherapy. Moving into this future will require the development of feedback and training systems that are based on the automated scoring of performance tasks like the MCO-PT. However, the availability of methods to reliably scale up the assessment of therapy-relevant behaviors, including the vital but potentially difficult to practice MCO, is a crucial developmental step. This study provides hope that the future of psychotherapy may involve opportunities for therapists to practice specific therapy skills and thereby, improve our clinical capacities and ultimately our clients' outcomes in treatment.

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

Allen, J. J., Parker, A., & Ogles, B. M. (2023). A review of the facilitative interpersonal skills performance task and rating method. *Clinical Psychology: Science and Practice*. Advance online publication. https://doi.org/10.1037/cps0000187

American Psychological Association. (2003). Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. *American Psychologist*, *58*(5), 377–402. https://doi.org/10.1037/0003-066X.58.5.377

Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology*, *84*(1), 57–66. https://doi.org/10.1037/ccp0000060

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology*, *65*(7), 755–768. https://doi.org/10.1002/jclp.20583

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*, *9*(1), Article 49. https://doi.org/10.1186/1748-5908-9-49

Atzil-Slonim, D., Juravski, D., Bar-Kalifa, E., Gilboa-Schechtman, E., Tuval-Mashiach, R., Shapira, N., & Goldberg, Y. (2021). Using topic models to identify clients' functioning levels and alliance ruptures in psychotherapy. *Psychotherapy*, *58*(2), 324–339. https://doi.org/10.1037/pst0000362

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Coppersmith, D. D., Dempsey, W., Kleiman, E. M., Bentley, K. H., Murphy, S. A., & Nock, M. K. (2022). Just-in-time adaptive interventions for suicide prevention: Promise, challenges, and future directions. *Psychiatry*, *85*(4), 317–333. https://doi.org/10.1080/00332747.2022.2092828

D'Andrea, M., Daniels, J., & Heck, R. (1991). Evaluating the impact of multicultural counseling training. *Journal of Counseling and Development*, *70*(1), 143–150. https://doi.org/10.1002/j.1556-6676.1991.tb01576.x

Davis, D. E., DeBlaere, C., Brubaker, K., Owen, J., Jordan, T., II, Hook, J., & Van Tongeren, D. (2016). Microaggressions and perceptions of cultural humility in counseling. *Journal of Counseling and Development*, *94*(4), 483–493. https://doi.org/10.1002/jcad.12107

Davis, D. E., DeBlaere, C., Owen, J., Hook, J. N., Rivera, D. P., Choe, E., Van Tongeren, D. R., Worthington, E. L., & Placeres, V. (2018). The multicultural orientation framework: A narrative review. *Psychotherapy*, *55*(1), 89–100. https://doi.org/10.1037/pst0000160

DeBlaere, C., Zelaya, D. G., Bowie, J. A., Chadwick, C. N., Davis, D. E., Hook, J. N., & Owen, J. (2023). Multiple microaggressions and therapy outcomes: The indirect effects of cultural humility and working alliance with black, indigenous, women of color clients. *Professional Psychology, Research and Practice*, *54*(2), 115–124. https://doi.org/10.1037/pro0000497

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. Arxiv. https://doi.org/10.48550/arXiv.1810.04805

Drinane, J. M., Owen, J., Adelson, J. L., & Rodolfa, E. (2016). Multicultural competencies: What are we measuring? *Psychotherapy Research*, *26*(3), 342–351. https://doi.org/10.1080/10503307.2014.983581

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Ehie, O., Muse, I., Hill, L., & Bastien, A. (2021). Professionalism: Microaggression in the healthcare setting. *Current Opinion in Anaesthesiology*, *34*(2), 131–136. https://doi.org/10.1097/ACO.0000000000000966

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*(1), 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Flemotomos, N., Martinez, V. R., Chen, Z., Creed, T. A., Atkins, D. C., & Narayanan, S. (2021). Automated quality assessment of cognitive behavioral therapy sessions through highly contextualized language representations. *PLOS ONE*, *16*(10), Article e0258639. https://doi.org/10.1371/journal.pone.0258639

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*(2), 690–711. https://doi.org/10.3758/s13428-021-01623-4

Freetly Porter, E., Owen, J., Agorsor, C., Kivlighan, M., Rousmaniere, T., Narvaez, C., & Heshmati, S. (2023). *Multicultural orientation deliberate experiencing training: Pilot study* [Manuscript in preparation].

Goldberg, S. B., Babins-Wagner, R., Imel, Z. E., Caperton, D. D., Weitzman, L. M., & Wampold, B. E. (2023). Threat alert: The effect of outliers on the

alliance-outcome correlation. *Journal of Counseling Psychology*, *70*(1), 81–89. https://doi.org/10.1037/cou0000638

Goldberg, S. B., Baldwin, S. A., Merced, K., Caperton, D. D., Imel, Z. E., Atkins, D. C., & Creed, T. (2020). The structure of competence: Evaluating the factor structure of the Cognitive Therapy Rating Scale. *Behavior Therapy*, *51*(1), 113–122. https://doi.org/10.1016/j.beth.2019.05.008

Goldberg, S. B., Flemotomos, N., Martinez, V. R., Tanana, M. J., Kuo, P. B., Pace, B. T., Villatte, J. L., Georgiou, P. G., Van Epps, J., Imel, Z. E., Narayanan, S. S., & Atkins, D. C. (2020). Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *Journal of Counseling Psychology*, *67*(4), 438–448. https://doi.org/10.1037/cou0000382

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, *63*(1), 1–11. https://doi.org/10.1037/cou0000131

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

Hargons, C., Mosley, D., Falconer, J., Faloughi, R., Singh, A., Stevens-Watkins, D., & Cokley, K. (2017). Black lives matter: A call to action for counseling psychology leaders. *The Counseling Psychologist*, *45*(6), 873–901. https://doi.org/10.1177/0011000017733048

Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: A systematic review. *Psychotherapy Research*, *30*(4), 417–432. https://doi.org/10.1080/10503307.2019.1620366

Hook, J. N., Davis, D. E., Owen, J., Worthington, E. L., Jr., & Utsey, S. O. (2013). Cultural humility: Measuring openness to culturally diverse clients. *Journal of Counseling Psychology*, *60*(3), 353–366. https://doi.org/10.1037/a0032595

Hook, J. N., Farrell, J. E., Davis, D. E., DeBlaere, C., Van Tongeren, D. R., & Utsey, S. O. (2016). Cultural humility and racial microaggressions in counseling. *Journal of Counseling Psychology*, *63*(3), 269–277. https://doi.org/10.1037/cou0000114

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kendi, I. X. (2019). *How to be an anti-racist*. One World.

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. ArXiv. https://doi.org/10.48550/arXiv.1412.6980

Kostick-Quenet, K. M., Cohen, I. G., Gerke, S., Lo, B., Antaki, J., Movahedi, F., Njah, H., Schoen, L., Estep, J. E., & Blumenthal-Barby, J. S. (2022). Mitigating racial bias in machine learning. *The Journal of Law, Medicine & Ethics*, *50*(1), 92–100. https://doi.org/10.1017/jme.2022.13

Kuo, P. B., Tanana, M. J., Goldberg, S. B., Caperton, D. D., Narayanan, S., Atkins, D. C., & Imel, Z. E. (2023). Machine learning-based prediction of client distress from session recordings. *Clinical Psychological Science*. Advance online publication. https://doi.org/10.1177/21677026231172694

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. ArXiv. https://doi.org/10.48550/arXiv.1907.11692

Open AI. (2022). *ChatGPT: Optimizing language models for dialogue*. Retrieved April 25, 2023, from https://openai.com/blog/chatgpt/

Owen, J. (2013). Early career perspectives on psychotherapy research and practice: Psychotherapist effects, multicultural orientation, and couple interventions. *Psychotherapy*, *50*(4), 496–502. https://doi.org/10.1037/a0034617

Owen, J., Imel, Z., Tao, K. W., Wampold, B., Smith, A., & Rodolfa, E. (2011). Cultural ruptures in short-term therapy: Working alliance as a mediator between clients' perceptions of microaggressions and therapy outcomes. *Counselling & Psychotherapy Research*, *11*(3), 204–212. https://doi.org/10.1080/14733145.2010.491551

Owen, J., Leach, M. M., Wampold, B., & Rodolfa, E. (2011). Client and therapist variability in clients' perceptions of their therapists' multicultural competencies. *Journal of Counseling Psychology*, *58*(1), 1–9. https://doi.org/10.1037/a0021496

Owen, J., Tao, K. W., & Drinane, J. (2019). Microaggressions: Clinical impact and psychological harm. In G. C. Torino, D. P. Rivera, C. M. Capodilupo, K. L. Nadal, & D. W. Sue (Eds.), *Microaggression theory: Influence and implications* (pp. 67–85). Wiley.

Owen, J. J., Tao, K., Leach, M. M., & Rodolfa, E. (2011). Clients' perceptions of their psychotherapists' multicultural orientation. *Psychotherapy*, *48*(3), 274–282. https://doi.org/10.1037/a0022065

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). *Automatic differentiation in pytorch*. https://openreview.net/forum?id=BJJsrmfCZ

Patel, S. B., & Lam, K. (2023). ChatGPT: The future of discharge summaries? *The Lancet Digital Health*, *5*(3), e107–e108. https://doi.org/10.1016/S2589-7500(23)00021-3

Pérez-Rojas, A. E., Bartholomew, T. T., Lockard, A. J., & González, J. M. (2019). Development and initial validation of the Therapist Cultural Comfort Scale. *Journal of Counseling Psychology*, *66*(5), 534–549. https://doi.org/10.1037/cou0000344

Pierce, C. M., Carew, J. V., Pierce-Gonzalez, D., & Wills, D. (1977). An experiment in racism: TV commercials. *Education and Urban Society*, *10*(1), 61–87. https://doi.org/10.1177/001312457701000105

Pine, G. J. (1972). Counseling minority groups: A review of the literature. *The National Catholic Guidance Conference Journal*, *17*(1), 35–44. https://doi.org/10.1002/j.2164-5183.1972.tb00209.x

Roberts, S. O., & Rizzo, M. T. (2021). The psychology of American racism. *American Psychologist*, *76*(3), 475–487. https://doi.org/10.1037/amp0000642

Rousmaniere, T. (2019). *Mastering the inner skills of psychotherapy: A deliberate practice manual*. Gold Lantern Books.

Rousmaniere, T., Goodyear, R. K., Miller, S. D., & Wampold, B. E. (Eds.). (2017). *The cycle of excellence: Using deliberate practice to improve supervision and training*. Wiley. https://doi.org/10.1002/9781119165590

Schöttke, H., Flückiger, C., Goldberg, S. B., Eversmann, J., & Lange, J. (2017). Predicting psychotherapy outcome based on therapist interpersonal skills: A five-year longitudinal study of a therapist assessment protocol. *Psychotherapy Research*, *27*(6), 642–652. https://doi.org/10.1080/10503307.2015.1125546

Schwartzman, C. M. (2022). *Therapist facilitative interpersonal skills in simulated text-based telepsychotherapy with cultural minority clients* [Doctoral dissertation, State University of New York at Albany]. ProQuest Dissertations & Theses Global.

Sennrich, R., Haddow, B., & Birch, A. (2015). *Neural machine translation of rare words with subword units*. Arxiv. https://doi.org/10.48550/arXiv.1508.07909

Soto, A., Smith, T. B., Griner, D., Domenech Rodríguez, M., & Bernal, G. (2018). Cultural adaptations and therapist multicultural competence: Two

meta-analytic reviews. *Journal of Clinical Psychology*, *74*(11), 1907–1923. https://doi.org/10.1002/jclp.22679

Stewart, S. H., Drinane, J. M., Owen, J., & Dumas, D. (2023). Psychotherapy training via a task-based assessment of the multicultural orientation framework: A pilot study. *Counselling & Psychotherapy Research*. Advance online publication. https://doi.org/10.1002/capr.12629

Tanana, M., Hallgren, K. A., Imel, Z. E., Atkins, D. C., & Srikumar, V. (2016). A comparison of natural language processing methods for automated coding of motivational interviewing. *Journal of Substance Abuse Treatment*, *65*, 43–50. https://doi.org/10.1016/j.jsat.2016.01.006

Tao, K. W., Owen, J., Pace, B. T., & Imel, Z. E. (2015). A meta-analysis of multicultural competencies and psychotherapy process and outcome. *Journal of Counseling Psychology*, *62*(3), 337–350. https://doi.org/10.1037/cou0000086

Tracey, T. J. G. (2016). A note on socially desirable responding. *Journal of Counseling Psychology*, *63*(2), 224–232. https://doi.org/10.1037/cou0000135

Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist*, *69*(3), 218–229. https://doi.org/10.1037/a0035099

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2019). *Huggingface's transformers: State-of-the-art natural language processing*. Arxiv. https://doi.org/10.48550/arXiv.1910.03771

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

Zhang, H., Watkins, C. E., Jr., Hook, J. N., Hodge, A. S., Davis, C. W., Norton, J., Wilcox, M. M., Davis, D. E., & Owen, J. (2022). Cultural humility in psychotherapy and clinical supervision: A research review. *Counselling & Psychotherapy Research*, *22*(3), 548–557. https://doi.org/10.1002/capr.12481

Received July 1, 2023
Revision received November 7, 2023
Accepted November 9, 2023 ■