Journal of Consulting and Clinical Psychology

Therapist–Client Agreement in Assessments of Clients' Functioning

Eran Bar-Kalifa, Dana Atzil-Slonim, Eshkol Rafaeli, Tuvia Peri, Julian Rubel, and Wolfgang Lutz

Online First Publication, October 24, 2016. http://dx.doi.org/10.1037/ccp0000157

Journal of Consulting and Clinical Psychology

0022-006X/16/$12.00 http://dx.doi.org/10.1037/ccp0000157

Therapist–Client Agreement in Assessments of Clients' Functioning

Eran Bar-Kalifa, Dana Atzil-Slonim, Eshkol Rafaeli, and Tuvia Peri
Bar-Ilan University

Julian Rubel and Wolfgang Lutz
University of Trier

Objective: We aimed to assess the extent to which therapists' reports of client functioning track their clients' changing experience of their own functioning from session to session (*temporal congruence*) as well as the extent to which they over- or underestimate their clients' functioning (*level* or *directional bias*) and to examine whether these indices predict treatment outcomes. ***Method:*** The participants included 384 clients who were treated by 77 therapists. Both clients and therapists rated the clients' functioning each session. The clients also completed pre- and posttreatment outcome measures. ***Results:*** Using multilevel modeling, we found that therapists' reports regarding their clients' functioning tended to be temporally congruent from session to session with their clients' reported functioning. In addition, on average, therapists did not show a level bias (i.e., did not over- or underestimate their clients' functioning). Finally, temporal congruence (but not level bias) predicted better treatment outcomes. ***Conclusion:*** These findings highlight the importance of tracking clients' fluctuating symptoms over time. Thus, we discuss their implication for the policy and practice of providing session-by-session feedback to therapists.

What is the public health significance of this article?
This study found that therapists' reports regarding their clients' functioning are temporally congruent with their clients' reports over time. It also suggests that such temporal congruence facilitates clients' progress in therapy.

Keywords: congruence, outcome, between- and within-dyad analysis, truth and bias model

The last two decades have seen a surge of research examining the extent to which providing clinicians with routine feedback regarding their clients' functioning improves therapeutic outcome (e.g., Hatfield & Ogles, 2006; Lambert, 2007; Shimokawa, Lambert, & Smart, 2010). The rationale behind providing clinicians with such feedback seems to begin with the premise that accurate knowledge of clients' level of symptomatology or functioning is a key factor that enables therapists to make appropriate treatment decisions (e.g., deciding whether a specific intervention is effective), which in turn can have a substantial beneficial impact on clients' overall therapeutic outcome (see Lambert & Shimokawa, 2011, for similar reasoning).

Adequate measurement of clients' functioning requires information from multiple sources (e.g., reports from the clients themselves, from trained judges, and from significant others, coupled with objective markers of daily functioning if available; Lambert, Hansen, & Finch, 2001). However, if one accepts clients' reports of their symptoms or functioning as one important criterion, one should expect that agreement between therapists' and clients' perceptions of clients' functioning would better equip therapists to take appropriate therapeutic action, as has been shown with regard to convergence between therapist and clients regarding other important processes, such as alliance (e.g., Laws et al., 2016).

It is interesting that few studies to date have empirically examined (a) whether this agreement actually exists and (b) whether it benefits clients. A frequent finding of these few studies is a moderate agreement between clients' and therapists' perceptions of treatment outcomes (e.g., Caplan, 1983; Holmqvist, Philips, & Mellor-Clark, 2016; Lampropoulos, 2010; Rosenblatt & Rosenblatt, 2002). Studies examining the *benefits* of therapist–client agreement have yielded mixed results. In one study, agreement regarding target complaints positively predicted treatment outcomes (Busseri & Tyler, 2004). Similarly, greater disagreement in clients' and therapists' perceptions of the severity of the clients' presenting problems was found to be associated with reduced odds of mutually agreed-upon termination (Corning, Malofeeva, & Bucchianeri, 2007). In another study, greater agreement was found between therapists and clients who completed treatment versus clients who unilaterally withdrew from treatment (Lampropoulos, 2010), although this difference failed to reach significance. By contrast, a more recent study found no association between client–

Eran Bar-Kalifa, Dana Atzil-Slonim, Eshkol Rafaeli, and Tuvia Peri, Department of Psychology, Bar-Ilan University; Julian Rubel and Wolfgang Lutz, Department of Psychology, University of Trier.

The first two authors contributed equally to this article.

This research was supported by the German Research Foundation (DFG, LU 660/8-1 and LU 660/10-1 to Wolfgang Lutz).

Correspondence concerning this article should be addressed to Eran Bar-Kalifa, Psychology Department, Bar-Ilan University, Ramat-Gan, Israel. E-mail: eran.barkalifa@gmail.com

therapist agreement regarding symptoms and treatment outcomes (Holmqvist et al., 2016).

The dearth of studies on this subject may stem from clinical investigators' tendency to use different measures to assess clients' and therapists' perspectives regarding clients' functioning or symptoms, making it difficult to get an exact comparison between the two parties' ratings. Moreover, studies that have examined agreement have rarely estimated it by comparing repeated reports of symptoms or functioning from both parties, and even those that did do so (e.g., Busseri & Tyler, 2004; Holmqvist et al., 2016) have tended to rely on a limited number of sessions (typically two or three), reflecting phases within therapy rather than session-to-session changes. Consequently, these studies are informative regarding the issue of dyadic differences in broad agreement but less informative regarding agreement across sessions within dyads.

A related limitation of the extant literature stems from the operationalization of agreement. The construct has typically been examined through correlations between clients' and therapists' ratings of outcomes (e.g., Lampropoulos, 2010) or observers' ratings of agreement (e.g., Busseri & Tyler, 2004). These help demonstrate the existence of broad agreement but obscure two important parameters of agreement, namely (a) temporal congruence and (b) directional (or level) bias.

According to Fletcher and Kerr (2010) *temporal congruence* (aka tracking accuracy) can be defined as the correlation between a set of judgments over time (e.g., the correlation between clients' and therapists' ratings of functioning as they cofluctuate over time); *directional bias* (aka level bias) can be defined as the differences in the mean levels of a specific judgment across a sample compared with a benchmark (e.g., the degree to which therapists over- or underestimate their clients' functioning compared to the clients' ratings). The distinction between these indices of agreement has received considerable theoretical attention over the years (e.g., Cronbach, 1955; Funder & Colvin, 1997; Kenny & Albright, 1987). However, its importance has largely gone unrecognized in the psychotherapy literature. Thus, several important empirical questions remain unanswered.

First, psychotherapy researchers have yet to use a within-dyad or session-by-session methodology to examine temporal congruence. Should one expect such agreement, on average, to exist, as has been found at the between-dyads level (e.g., Holmqvist et al., 2016)? Moreover, will dyad-level differences in session-by-session temporal congruence predict treatment outcomes?

Second, to the best of our knowledge, no study has examined the direction of the differences (i.e., the possible directional bias) between therapists' and clients' perspectives. Such directional biases (i.e., over- or underestimation) may be meaningful and influence therapists' ability to take appropriate action when needed. Should one expect therapists, on average, to over- or underestimate their clients' functioning session to session? Furthermore, will dyad-level differences in directional bias predict treatment outcomes?

Third, to the best of our knowledge, no study has explored the association between temporal congruence and level bias in client–therapist ratings of treatment outcomes. Will within-dyad temporal congruence decrease as directional bias increases, or, instead, will the same therapists be both directionally biased and temporally congruent in their tracking of their clients' reported functioning?

A recent statistical innovation in judgment research—the truth and bias model developed by West and Kenny (2011)—is particularly appropriate for exploring these three questions because it allows a simultaneous within-dyad examination of (a) temporal congruence, (b) directional bias, and (c) the association between these two indices. Moreover, it can be used to obtain dyad-level differences in the two indices, which can then be used to predict treatment outcomes.

The truth and bias model, which was introduced in research on close relationships (Overall, Fletcher, & Kenny, 2012), has recently been utilized in psychotherapy research. Using this model, Atzil-Slonim et al. (2015) examined the temporal congruence and directional bias between clients' and therapists' session-by-session ratings of the therapeutic alliance and the association between these two indices. Their results indicated that therapists tended to accurately track changes in the therapeutic alliance but that their reports underestimated this alliance compared with their clients' reports. Of importance, temporal congruence and underestimation were positively associated. On the basis of similar patterns that have been found in close relationship research (e.g., Overall et al., 2012), Atzil-Slonim et al. interpreted the findings as reflecting a "better safe than sorry" pattern and argued that this pattern is adaptive because it helps therapists simultaneously avoid the risk of missing potential ruptures due to an overestimation of the alliance and remain attuned to their clients' changing experiences.

The Present Study

The present study aimed to examine temporal congruence, level bias, and the association between these two indices with regard to clients' and therapists' ratings of the clients' functioning. It also aimed to determine whether dyad-level differences in these two indices are associated with treatment outcomes. In each session, both the client and the therapist completed parallel measures to evaluate clients' functioning. These measures allowed us to fully utilize the multilevel model (MLM) version of the truth and bias model. Five specific hypotheses guided our work.

On the basis of the premise that therapists are motivated to be cognizant of their clients' fluctuating levels of functioning, we expected to find significant session-level temporal congruence (as suggested in studies showing dyad-level correlations between the two parties' reports; Hypothesis 1). On the basis of findings from close relationship research (e.g., Overall et al., 2012) and findings regarding congruence in therapeutic alliance ratings (Atzil-Slonim et al., 2015), we expected to find that therapists significantly underestimate their clients' functioning (Hypothesis 2). We also expected that therapists who were more negatively biased would demonstrate greater temporal congruence (Hypothesis 3). Finally, we expected greater temporal congruence (Hypothesis 4) and greater negative direction bias (Hypothesis 5) to predict better treatment outcomes.

Method

Participants and Treatment

Data collection was part of the routine outcome monitoring system implemented at the Trier University Outpatient Clinic in southwest Germany. All patients who started treatment were con-

sidered eligible for this study, as long as they provided at least nine session-level reports of functioning (to ensure sufficient within-client measurement points) and completed the Outcome Questionnaire–30 (OQ-30; Ellsworth, Lambert, & Johnson, 2006) and the Beck Depression Inventory–II (BDI-II; Beck, Steer, & Brown, 1996) at the beginning and end of the treatment. Of the 456 patients who entered treatment, 66 were excluded because of insufficient number of sessions, and six were excluded due to missing OQ-30 and/or BDI-II posttreatment scores (total $N_{excluded} = 72$). The excluded group had marginally lower pretreatment OQ-30 than did the included group ($M_{excluded} = 53.08$, $SD = 14.95$; $M_{included} = 56.9$, $SD = 15.3$), $t(454) = -1.950$, $p = .052$, but did not differ in regard to their pretreatment BDI-II scores ($M_{excluded} = 23.36$, $SD = 11.72$; $M_{included} = 23.87$, $SD = 11.04$), $t(454) = -.352$, $p = .725$. Thus, the analyses were based on a sample of 384 clients treated by 77 therapists. On average, clients received 36.19 treatment sessions ($SD = 16.84$). All the clients were older than 12 years ($M = 37.55$, $SD = 12.90$), and most clients were female (66.1%). Psychiatric diagnoses were based on the Structured Clinical Interview for Axis I *DSM–IV* Disorders—Patient Edition (SCID-I/P; First, Spitzer, Gibbon, & Williams, 1995), which was conducted before the actual therapy by independent clinicians who were trained in administering the SCID-I/P. Most of the clients were diagnosed with a depressive disorder (60.1%) or an anxiety disorder (16.1%) as the primary diagnosis. Additional primary diagnoses were acute stress and adjustment disorders (12%), eating disorders (2.1%), and other disorders (9.5%). To diagnose personality disorders, we adopted the International Diagnostic Checklist for Personality Disorders (Bronisch, Hiller, Mombour, & Zaudig, 1996). This checklist identified 77 of the 384 clients as having a personality disorder (20.1%). All the therapists participated in a 3- to 5-year postgraduate training program with a cognitive–behavioral therapy focus.

Instruments and Data Collection

Outcome Questionnaire–30 (OQ-30; Ellsworth et al., 2006). The OQ-30 was administered before and after treatment. This 30-item self-report measure is designed to assess client outcomes during the course of therapy. The OQ-30 has three primary dimensions: (a) subjective discomfort, (b) interpersonal relationships, and (c) social role performance. All 30 items can be aggregated to create a total score. Total scores can range from 0 to 120, with higher scores reflecting poorer psychological functioning. The OQ-30 is a short form of the OQ-45. It comprises the 30 items that are most sensitive to client change, and it has demonstrated high levels of congruence with the OQ-45 in the measurement of client outcomes (Ellsworth et al., 2006; Vermeersch, Lambert, & Burlingame, 2000; Vermeersch et al., 2004). The OQ-30 showed adequate internal consistency in our sample (pretreatment: $\alpha = .89$; posttreatment: $\alpha = .95$).

Beck Depression Inventory–II; Beck et al., 1996. The BDI-II was administered before and after treatment. This 21-item self-report measure asks respondents to rate the severity of their depressive symptoms over the previous 2 weeks on a variable Likert scale (i.e., 19 items use a 4-point scale, and two items use a 7-point scale). Individual item scores are summed to create a total severity score that ranges from 0 to 63. Total scores can be used to categorize respondents according to their depressive severity using the following ranges: 0 to 13 (minimal), 14 to 19 (mild), 20 to 28 (moderate), and >28 (severe; Beck et al., 1996). Analyses have revealed high internal consistency ($\alpha = .93$) and significant ($p < .01$) intercorrelations between the BDI-II total scale and the Behavior and Symptom Identification Scale–24's Depression/Functioning ($r = .79$) and Overall ($r = .82$) subscales (Subica et al., 2014). The BDI-II showed good internal consistency in our sample (pretreatment: $\alpha = .90$; posttreatment: $\alpha = .93$).

Functioning. Client functioning was assessed from the client's (functioning-C) and the therapist's (functioning-T) perspectives using two parallel questions. Before each session, the clients were asked the following question: "Over the past seven days, how well do you feel that you are getting along emotionally and psychologically?" After each session, the therapists were asked the following question: "How well is your patient getting along emotionally and psychologically?" The clients answered this question using the following 5-point Likert scale: 1 (*Quite poorly*), 2 (*Fairly poorly*), 3 (*So-so*), 4 (*Fairly well*), and 5 (*Quite well*). The therapists used the same scale, although they had an additional category: 6 (*Very well*). To create strictly parallel scales for clients and therapists, we merged the last two categories of the therapists' item into one category. The correlation between the therapists' original and adapted scales was extremely high ($r = .98$).

These questions were extracted from the Compass tracking system (Howard, Brill, Lueger, O'Mahoney, & Grissom, 1995), which was designed to assess treatment progress as part of outcome monitoring and feedback systems (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Lutz, Rafaeli, Howard, & Martinovich, 2002). The correlations between the function-C score and the 30-item short form of the OQ-45 for the entire sample of clients who were treated at the outpatient clinic from which the data were obtained were $r = -.52$ ($N = 1{,}194$) after Session 5, $r = -.54$ ($N = 1{,}077$) after Session 10, and $r = -.56$ ($N = 961$) after Session 15. Similarly, the correlations between the function-T score and the Global Assessment Scale, a measure of overall functioning (Endicott, Spitzer, Fleiss, & Cohen, 1976) were $r = .51$ ($N = 1{,}387$) after Session 5, $r = .56$ ($N = 1{,}190$) after Session 10, and $r = .52$ ($N = 1{,}074$) after Session 15. These moderate associations provide some evidence for the validity of these two questions as measures of client functioning.

Data collection and therapist feedback. Background measures (including BDI-II and OQ-30) were collected before the first session and following the final session. Prior to each (typically weekly) session, clients completed a brief report of symptoms along with the functioning item described earlier. Following each session, therapists completed a brief postsession report that included the therapist version of the functioning item. In general, therapists had full access to their clients' self-report data and were encouraged to use this feedback; however, they did not have access to their clients' presession report until they themselves had completed their own (postsession) report for that session.

Data Analytic Strategy

The data set had a hierarchical structure, with session ratings nested within clients and clients nested within therapists. Thus, we used a three-level MLM (Raudenbush & Bryk, 2002), with sessions at Level 1, clients at Level 2, and therapists at Level 3. This approach partitions the total variability into functioning ratings for

session *s* of client *c,* who was treated by therapist *t*, into three components: variance within client or therapist ratings at Level 1, between clients at Level 2, and between therapists at Level 3.

To test our primary hypotheses regarding temporal congruence and directional bias in therapists' judgments regarding their clients' functioning on a session-by-session basis, we used West and Kenny's (2011) truth and bias model. In these analyses, the therapists' reports of functioning constituted the *judgment,* which served as the dependent variable. The judgment was predicted by the *truth,*[1] which was the clients' reports of their functioning. The slope coefficient of the truth (i.e., the *truth force*) represented the degree to which the therapists were temporally congruent in judging their clients' functioning. As West and Kenny suggested (pp. 374–375), we centered the judgment and truth parameters on each person's mean truth score (i.e., the client's mean functioning) across all sessions, which allowed us to remove broad individual differences when examining within-subject fluctuations. In turn, the intercept estimate represented the *directional bias* (i.e., the extent to which therapists overestimate [in cases of positive intercepts] or underestimate [in cases of negative intercepts] their clients' functioning). We ran a three-level model in which (a) the directional bias (i.e., the intercept), (b) the truth force, and (c) the covariation between the two were treated as random at both Level 2 (i.e., the client level) and Level 3 (i.e., the therapist level).

The model's multilevel equation was as follows:

$$\text{Judgment}_{sct} = (\gamma_{000} + u_{00t} + r_{0ct}) + (\gamma_{100} + u_{10t} + r_{1ct}) * \text{Truth}_{sct} + e_{sct},$$

where the judgment for session *s* of client *c,* who was treated by therapist *t*, was predicted by the sample average (i.e., fixed) directional bias (i.e., the intercept; γ_{000}) and truth force (i.e., the slope; γ_{100}) multiplied by this session's truth variable (i.e., Truth_{sct}), plus this therapist's deviation for this particular client from the fixed directional bias and truth force (i.e., random effects at Level 3; u_{00t}, u_{10t}), plus this client's deviation from the average directional bias and truth force of the other clients who were treated by the same therapist (i.e., random effects at Level 2; r_{0ct}, r_{1ct}), plus a residual term quantifying the session's deviation from these effects (i.e., random effect at Level 1; e_{sct}).

Results

Table 1 displays the descriptive statistics for the study variables. Table 2 presents the results of the truth and bias analysis. The fixed effect for the truth variable was positive and significant, indicating that the therapists tended to be temporally congruent with their clients in their assessment of functioning, as predicted (Hypothesis 1). The random effect of the truth variable was also significant at both Level 2 and Level 3, indicating significant between-clients and between-therapists variabilities in the extent to which therapists were temporally congruent with their clients in their reports of functioning. Despite this variability, the client-specific truth variable for all therapists was positive, indicating that congruence was indeed the standard.

To estimate the explained variable in our model, we compared it to an unconditional three-level model in which judgment was the outcome (with no predictors other than the intercept; see Snijders & Bosker, 1999). As expected, our model that included only a Level 1 predictor (i.e., the truth variable) explained a considerable amount of the Level 1 variance (reducing it from .452 to .341, a reduction of 24.6%) but none of the Level 2 or Level 3 variance.

To rule out the possibility that the positive truth force reflected a positive collinear trend of the effect of treatment time (i.e., both clients' and therapists' ratings increased with time; see Bolger & Laurenceau, 2013), we reran the model with *time* (coded as 0 during the middle session) and *time by truth force* terms included. In this model, the main effect of truth force remained significant (estimate = .35, SE = .02, $p < .001$); however, we also found that the Time × Truth Force interaction term was significant (estimate = .004, SE = .0004, $p < .0001$), indicating that, as treatment progressed, therapists tended to have greater truth force (i.e., greater temporal congruence).

The fixed directional bias was negative but not significant, indicating that, in contrast to our prediction (Hypothesis 2), on average, therapists did not tend to underestimate their clients' functioning. However, the random effect of the directional bias was significant at both Level 2 and Level 3, indicating significant between-clients and between-therapists variabilities in terms of the extent to which therapists tended to under- or overestimate their clients' functioning. Specifically, whereas the therapist had a negative directional bias (i.e., underestimated the clients' functioning) in approximately half (52.6%) of the therapeutic dyads, the therapist had a positive directional bias in the other half. Finally, in contrast to our prediction (Hypothesis 3), we failed to find a negative association between directional biases and truth forces at either the client or the therapist level.

To test whether truth forces and directional biases predicted treatment outcomes (Hypotheses 4 and 5, respectively), we first obtained (for each client and for both the directional bias and the truth force) the empirical Bayes estimates (Raudenbush & Bryk, 2002) that were computed from the multilevel truth and bias model described earlier. Then, we used both of these estimates to predict posttreatment outcomes (i.e., BDI-II and OQ-30) while adjusting for the pretreatment levels of these variables. Although the outcome data were hierarchically nested (with two levels: clients within therapists), an unconditional two-level model of both outcomes revealed that the Level 2 (therapist level) variance was not significant (BDI-II: estimate = .003, SE = .008, p = .347; OQ-30: estimate = .003, SE = .013, p = .408) or substantial (BDI-II: intraclass correlation [ICC] = .01; OQ-30: ICC = 0.01). Therefore, we opted to use ordinary least squares regressions (one for each outcome measure), with clients as the (single) level of analysis. As predicted (Hypothesis 4), the truth force was negatively associated with both BDI-II ($b = -.82$, SE = .14, $\beta = -.26$, $p < .001$) and OQ-30 ($b = -1.00$, SE = .19, $\beta = -.24$, $p < .001$): Greater temporal congruence predicted better treatment outcomes. In addition, in partial support of our prediction (Hypothesis 5), the directional bias was marginally associated with posttreatment OQ-30 (b = .10, SE = .06, β = .08, p = .083) but not with posttreatment BDI-II (b = .03, SE = .04, β = .03, p = .512).

[1] Of note, we opted to use the term *truth* to be consistent with the technical terms used in the truth and bias model to describe the model's parameters. We do not wish to imply that clients' reports (or therapists' reports) should be taken as "truth" regarding the objective levels of functioning.

Table 1
Mean, Standard Deviation, Ranges, and Inter-Correlations of Study Variables

Variable	*M*	*SD*	Range	Zero-order correlations[a] 2	3	4	5	6
1. Client reports of functioning	3.49	.92	1–5	.57	−.42	−.55	−.44	−.51
2. Therapist reports of functioning	3.59	.82	1–5	—	−.40	−.51	−.38	−.49
3. Pretreatment OQ-30	56.90	15.30	8–98		—	.39	.74	.36
4. Posttreatment OQ-30	34.31	18.29	1–90			—	.36	.82
5. Pretreatment BDI-II	23.87	11.04	0–55				—	.44
6. Posttreatment BDI-II	10.41	9.82	0–54					—

Note. OQ-30 = Outcome Questionnaire–30; BDI-II = Beck Depression Inventory–II.
[a] The zero-order correlations with the clients' and therapists' reports of functioning were computed using these variables' means (computed across all treatment sessions). The within-client zero-order correlations between these two variables (computed as the association between the variables after removing their averages) was $r = .45$.

We conducted an exploratory analysis to determine whether clients' pretreatment symptom severity had any impact on the extent to which the dyads would be in agreement regarding the clients' functioning. For example, it may be that the higher symptom levels are more overt and thus easier to follow; additionally, it may be that clients with higher levels of symptoms are also more interpersonally impaired and thus communicate their functioning levels less clearly, which in turn may hamper their therapists' ability to track this functioning. To test this idea, we ran another truth and bias model, in which both pretreatment OQ-30 and BDI-II were entered as linear and quadratic moderators of directional bias and temporal congruence. This analysis yielded little support for this idea. Specifically, no quadratic effect was found for either pretreatment OQ-30 or pretreatment BDI-II on either directional bias or temporal congruence. Similarly, neither pretreatment index had a linear effect on directional bias, and pretreatment OQ-30 had no linear effect on temporal congruence. The sole (marginally significant) finding was with pretreatment BDI-II, which moderated temporal congruence (estimate = .003, SE = .0017, $p = .0535$): Dyads in which the clients had higher pretreatment BDI-II scores had marginally higher temporal congruence.

Discussion

The current study examined therapist–client agreement in assessing clients' functioning, as reflected by two indices of agreement: (a) temporal congruence (i.e., the covariation between the two parties' assessments) and (b) directional bias (i.e., the difference in the levels between the two parties' assessments). It also examined the association between the two indices and the degree to which they predict treatment outcomes.

The results supported Hypothesis 1—namely, that therapists would be temporally congruent with their clients with regard to functioning. This finding is consistent with those of previous studies that have found a moderate association between clients' and therapists' ratings of the clients' outcomes (e.g., Lampropoulos, 2010). Of importance, previous studies have examined the association at the dyad level, whereas the current study is the first to show this association at the session level as well.

Contrary to Hypothesis 2, we failed to find a negative directional bias in the therapists' ratings—that is, on average, therapists did not tend to underestimate their clients' functioning. To our knowledge, no study to date has examined the possibility of directional biases in assessments of clients' functioning. An interesting perspective on these (null) results may be drawn from a recent meta-analysis (Sánchez-Meca, Rosa-Alcázar, Marín-Martínez, & Gómez-Conesa, 2010) that compared treatment outcome effect sizes on the basis of clients' and therapists' perspectives. Notably, the meta-analysis found certain domains of functioning for which therapists seemed to be less optimistic than were their clients (e.g., depression, global adjustment) and other domains of functioning for which these roles were reversed (e.g., anxiety; see also Kraus et al., 2016; Ogles, 2013). The current study used a very general item to assess functioning. With more domain-specific items, underestimation (but also overestimation) could have emerged.

We did not find the expected association between temporal congruence and directional bias (Hypothesis 3). In other words, the better-safe-than-sorry pattern found in a recent study examining congruence in client–therapist alliance ratings (Atzil-Slonim et al., 2015) and among romantic partners (Overall et al., 2012) was not evident in the client–therapist ratings of the client's functioning.

Table 2
Truth and Bias Model for Ratings of Functioning

Parameter estimate	Estimate (*SE*)	*p*	Effect size[a]
Fixed effects			
Intercept (directional bias) (γ_{000})	−.055 (.036)	.131	.04
Slope (truth force) (γ_{100})	.414 (.016)	<.001	.92
Random effects			
Level 1 (sessions)			
Residual	.341 (.004)	<.001	
Level 2 (clients)			
Intercept	.196 (.017)	<.001	
Slope	.029 (.004)	<.001	
Intercept/slope covariance	−.004 (.006)	.465	
Level 3 (therapists)			
Intercept	.047 (.017)	.003	
Slope	.007 (.003)	.009	
Intercept/slope covariance	.004 (.005)	.427	
Model summary			
−2 LogL (deviance)	28,193.5		
No. estimated parameters	9		

[a] Effect sizes were estimated with semipartial R^2 for linear mixed models (Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008).
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

This finding, together with the absence of an average directional bias (Hypothesis 2), may suggest that the vigilant approach of underestimation is less relevant for the evaluation of functioning than it is for the evaluation of relational factors such as alliance. Nonetheless, it is possible that a vigilant approach regarding relational factors may facilitate greater agreement regarding nonrelational factors, including the evaluation of functioning. For example, it may be that therapists who adopt the better-safe-than-sorry approach regarding the state of the therapeutic relationship would be more motivated to negotiate with their clients toward agreement regarding the clients' level of functioning. Future studies may wish to test this hypothesis by examining client–therapist agreement across multiple factors (e.g., alliance and functioning).

In support of Hypothesis 4, greater temporal congruence predicted better treatment outcomes. The literature regarding the association between client–therapist agreement concerning clients' symptoms or functioning and treatment outcomes has been mixed, with several studies showing a positive association (e.g., Busseri & Tyler, 2004; Corning et al., 2007; Lampropoulos, 2010) and others not showing such an association (e.g., Holmqvist et al., 2016). However, previous studies have often utilized nonparallel measures to assess clients' and therapists' perspectives and have rarely utilized intensive repeated reports. To our knowledge, the present study is the first to address these methodological considerations, which we believe allowed a more reliable and valid demonstration of the association between therapist–client agreement and treatment outcomes.

The association between temporal congruence and treatment outcomes is consistent with the idea that therapists who are aware of their clients' functioning are better equipped to help their clients. It also fits with the growing evidence that feedback provided to (or collected by) therapists regarding changes in the clients' level of functioning may be a powerful tool for enhancing psychotherapy outcomes (Krägeloh, Czuba, Billington, Kersten, & Siegert, 2015; Lambert & Shimokawa, 2011; Lutz, De Jong, & Rubel, 2015). Specifically, information regarding changes in clients' level of functioning may be crucial in helping therapists take appropriate action or reconsider the treatment formulation, which in turn may lead to better treatment outcomes (Castonguay, Barkham, Lutz, & McAleavey, 2013). It is interesting that all of our therapists received routine feedback regarding their clients' progress, and their assessments were probably influenced (i.e., biased) by feedback they had received earlier in the therapy (e.g., about the trajectory of their clients' functioning up to the previous session). Nonetheless, we found substantial variation in the levels of therapist–client agreement, which in turn were tied to treatment outcome.

Previous studies (e.g., Lutz et al., 2015) have found attitudes toward feedback and actual use of it to moderate the effects of feedback. Unfortunately, data about therapist attitudes and usage were not collected in the present study, but it certainly seems fruitful to explore the possibility that therapist–client agreement mediates this moderation—that is, that therapists who are enthusiastic about feedback may indeed have greater agreement with their clients and, in turn, will have better outcomes. Additionally, there is some evidence that feedback is particularly beneficial for clients who are at risk of treatment failure (e.g., Krägeloh et al., 2015). It will be important to determine whether the process by which feedback aids such treatment involves greater agreement (and especially, temporal congruence).

An alternative explanation for the association between temporal congruence and treatment outcomes is that clients who benefit more from treatment communicate their functioning more clearly to their therapists—that is, that the clients' improvement increases their readability, which in turn drives the dyad's congruence. To disentangle these alternative explanations, future studies should go a step further and examine the processes that occur within psychotherapy sessions (including the therapists' choice of interventions and the clients' clarity of communication), which may mediate the association between agreement and outcomes.

Hypothesis 5—namely, that the directional bias would be associated with treatment outcomes—was only partially supported. Specifically, directional bias was marginally associated with one measure of treatment outcomes (OQ-30), and it was not associated with the second measure (BDI-II). This result seems to suggest that directional bias is a much less robust predictor of treatment outcome than is temporal congruence. In other words, therapists' ability to be congruent with their clients with regard to *changes* in their clients' experienced functioning is seemingly more important than their ability to perceive the actual levels of their clients' experienced functioning.

Additional Future Directions

In this study, we employed both client and therapist reports as our solitary sources of data. This limits our ability to establish whether therapist–client agreement also involves accurate assessment of clients' subjective level of functioning or to establish whether clients and/or therapists are objectively accurate. Future studies could help disentangle this agreement or accuracy or truth question by obtaining therapists' perception of clients' own subjective assessments (e.g., "How does your patient *perceive* himself/herself as getting along emotionally and psychologically"), which would allow us to examine the effect of therapists' accuracy regarding their clients' subjective states; alternatively, studies may benefit from using objective indices (or judges) to assess functioning in-session or from using informers (e.g., friends, spouse) to obtain estimates of out-of-session functioning; these would allow one to examine the effect of therapists' accuracy regarding their clients' objective states.

Relatedly, in the current study we used a single item for assessing functioning; this choice alleviates reporters' burden but unfortunately may limit the variability of the assessed construct, which in turn may reduce the power to detect both directional bias (e.g., because the limited scale may pull for agreement) or temporal congruence (e.g., because the limited variability may restrict covariation). Still, it is important to note that even with this gross one-item index, we found significant between-therapists and between-clients variability in both temporal congruence and directional bias (though both might have been even greater if we had used a more sensitive index of functioning).

Such variability signals the need for future empirical efforts to identify relevant moderators, which may include characteristics of the idiosyncratic therapy process itself and the therapists' or the clients' pretreatment characteristics (e.g., DeRubeis, Gelfand, German, Fournier, & Forand, 2014). In the current study, we failed to find any robust moderation effect of clients' pretreatment func-

tioning on therapist–client agreement; however, future studies should try to examine other potential client or therapist factors such as attachment orientation (e.g., Mallinckrodt, 2000), interpersonal problems (e.g., Hersoug, Høglend, Havik, von der Lippe, & Monsen, 2009), and emotional regulation ability (Gratz & Tull, 2010). The inclusion of such moderators may also help in identifying those cases in which agreement facilitates treatment outcome and those in which it may actually hinder treatment outcome. For example, it has been suggested that people with attachment anxiety use complaints to elicit care (Stuart & Noyes, 1999); it may be that therapists who are able to empathically diverge from this often ruminative preoccupation with the complaints are in a better position to help them develop constructive means to elicit care. More generally, it may be that a certain level of incongruence might be useful at times to promote insight and change—and thus, that congruence does not necessarily exert only a linear effect. Of course, these speculations should be empirically examined.

Additionally, in the current study we obtained omnibus dyad-level indices of agreement (i.e., directional bias and temporal congruence that were estimated across all the sessions of each client) and then used these indices to predict treatment outcome. We adopted this approach because temporal congruence, the agreement index that proved to be predictive of treatment outcome, is definable only across time. However, a more granular analysis of therapist–client agreement and its immediate consequences (i.e., possible symptomatic relief in the next session) is a worthy target of examination for future research.

It is also important to note that the therapeutic approach of all the therapists in this study was cognitive–behavioral; such homogeneity limits the generalizability of the present findings. For example, it is possible that cognitive–behavioral therapy's focus on symptomatic change makes agreement in this domain a predictor of treatment outcomes. Indeed, with other therapeutic approaches, other domains of phenomenology (e.g., emotional experience: e.g., Greenberg & Paivio, 2003; or interpersonal experience: Klerman & Weissman, 1994) may play a more central role, and agreement regarding these domains may have a stronger predictive value. To test this idea, future studies should examine these processes for different therapeutic approaches, explore agreement in various domains, and consider various types of outcomes.

Summary and Clinical Implications

Clinical wisdom often assumes that therapists' understanding of clients' symptoms and functioning is a prerequisite for effective intervention. It is interesting that this wisdom has rarely been empirically tested. We aimed to assess the extent to which therapists' agreement with their clients regarding changes in their functioning from session to session (i.e., temporal congruence)—as well as the extent to which therapists over- or underestimate their clients' functioning (i.e., directional bias)—predicts treatment outcomes. We used session-by-session reports from both clients and their therapists, subjecting these reports to the tools offered by West and Kenny's (2011) truth and bias model. Our results found evidence of considerable therapist–client agreement (manifest in both temporal congruence and minimal directional bias). However, only temporal congruence proved to be associated with treatment outcomes.

Though further studies are needed to clarify the processes in question, these results highlight the greater importance of tracking clients' fluctuating levels of functioning over time—that is, their standing relative to themselves—even more than their absolute level of functioning. From a clinical standpoint, it may be less productive for therapists to try to achieve absolute agreement with their clients regarding their "objective" levels of functioning than to be aware of the trajectories of functioning as the treatment unfolds. Such awareness may be attained by the routine monitoring of clients' functioning, as well as by facilitating open communication between clients and therapists regarding functioning.

References

Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., Lutz, W., Rubel, J., Schiefele, A. K., & Peri, T. (2015). Therapeutic bond judgments: Congruence and incongruence. *Journal of Consulting and Clinical Psychology, 83,* 773–784. http://dx.doi.org/10.1037/ccp0000015

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.

Bronisch, T., Hiller, W., Mombour, W., & Zaudig, M. (1996). *International diagnostic checklists for personality disorders according to ICD-10 and DSM–IV—IDCL-P*. Seattle, WA: Hogrefe and Huber.

Busseri, M. A., & Tyler, J. D. (2004). Client-therapist agreement on target problems, working alliance, and counseling outcome. *Psychotherapy Research, 14,* 77–88. http://dx.doi.org/10.1093/ptr/kph005

Caplan, B. (1983). Staff and patient perception of patient mood. *Rehabilitation Psychology, 28,* 67–77. http://dx.doi.org/10.1037/h0091003

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. (2013). Practice-oriented research: Approaches and applications. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (6th ed., pp. 85–133). New York, NY: Wiley.

Corning, A. F., Malofeeva, E. V., & Bucchianeri, M. M. (2007). Predicting termination type from client-therapist agreement on the severity of the presenting problem. *Psychotherapy: Theory, Research, Practice, Training, 44,* 193–204. http://dx.doi.org/10.1037/0033-3204.44.2.193

Cronbach, L. J. (1955). Processes affecting scores on understanding of others and assumed similarity. *Psychological Bulletin, 52,* 177–193. http://dx.doi.org/10.1037/h0044919

DeRubeis, R. J., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014). Understanding processes of change: How some patients reveal more than others—and some groups of therapists less—about what matters in psychotherapy. *Psychotherapy Research, 24,* 419–428. http://dx.doi.org/10.1080/10503307.2013.838654

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine, 27,* 6137–6157. http://dx.doi.org/10.1002/sim.3429

Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A comparison of the Outcome Questionnaire-45 and Outcome Questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology & Psychotherapy, 13,* 380–391. http://dx.doi.org/10.1002/cpp.503

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33,* 766–771. http://dx.doi.org/10.1001/archpsyc.1976.01770060086012

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). *Structured Clinical Interview for Axis I DSM–IV Disorders—Patient ed. (SCID-I/P)*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute.

Fletcher, G. J. O., & Kerr, P. S. G. (2010). Through the eyes of love: Reality and illusion in intimate relationships. *Psychological Bulletin, 136,* 627–658. http://dx.doi.org/10.1037/a0019792

Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego, CA: Academic Press.

Gratz, K. L., & Tull, M. T. (2010). Emotion regulation as a mechanism of change in acceptance- and mindfulness-based treatments. In R. Baer (Ed.), *Assessing mindfulness and acceptance processes in clients: Illuminating the theory and process of change* (pp. 107–133). Oakland, CA: New Harbinger.

Greenberg, L. S., & Paivio, S. C. (2003). *Working with emotions in psychotherapy* (Vol. 13). New York, NY: Guilford Press.

Hatfield, D. R., & Ogles, B. M. (2006). The influence of outcome measures in assessing client change and treatment decisions. *Journal of Clinical Psychology, 62,* 325–337. http://dx.doi.org/10.1002/jclp

Hersoug, A. G., Høglend, P., Havik, O. E., von der Lippe, A., & Monsen, J. T. (2009). Pretreatment patient characteristics related to the level and development of working alliance in long-term psychotherapy. *Psychotherapy Research, 19,* 172–180. http://dx.doi.org/10.1080/10503300802657374

Holmqvist, R., Philips, B., & Mellor-Clark, J. (2016). Client and therapist agreement about the client's problems—Associations with treatment alliance and outcome. *Psychotherapy Research, 26,* 399–409. http://dx.doi.org/10.1080/10503307.2015.1013160

Howard, K. I., Brill, P. L., Lueger, R. J., O'Mahoney, M. T., & Grissom, G. R. (1995). *Integrating outpatient tracking assessment*. Philadelphia, PA: Compass Information Services, Inc.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51,* 1059–1064. http://dx.doi.org/10.1037/0003-066X.51.10.1059

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102,* 390–402. http://dx.doi.org/10.1037/0033-2909.102.3.390

Klerman, G. L., & Weissman, M. M. (1994). *Interpersonal psychotherapy of depression: A brief, focused, specific strategy*. Lanham, MD: Aronson.

Krägeloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome measures in mental health services: A scoping study and typology. *Psychiatric Services, 66,* 224–241. http://dx.doi.org/10.1176/appi.ps.201400141

Kraus, D. R., Bentley, J. H., Alexander, P. C., Boswell, J. F., Constantino, M. J., Baxter, E. E., & Castonguay, L. G. (2016). Predicting therapist effectiveness from their own practice-based evidence. *Journal of Consulting and Clinical Psychology, 84,* 473–483. http://dx.doi.org/10.1037/ccp0000083

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17,* 1–14. http://dx.doi.org/10.1080/10503300601032506

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172. http://dx.doi.org/10.1037/0022-006X.69.2.159

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48,* 72–79. http://dx.doi.org/10.1037/a0022238

Lampropoulos, G. K. (2010). Type of counseling termination and trainee therapist-client agreement about change. *Counselling Psychology Quarterly, 23,* 111–120. http://dx.doi.org/10.1080/09515071003721552

Laws, H. B. Constantino, M. J., Sayer, A., Klein, D. N., Kocsis, J. H., Manber, R., . . . Arnow, B. A. (2016). Convergence in patient-therapist therapeutic alliance ratings and its relation to outcome in chronic depression treatment. *Psychotherapy Research.* Advance online publication. http://dx.doi.org/10.1080/10503307.2015.1114687

Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research, 25,* 625–632. http://dx.doi.org/10.1080/10503307.2015.1079661

Lutz, W., Rafaeli, E., Howard, K. I., & Martinovich, Z. (2002). Adaptive modeling of progress in outpatient psychotherapy. *Psychotherapy Research, 12,* 427–443. http://dx.doi.org/10.1093/ptr/12.4.427

Mallinckrodt, B. (2000). Attachment, social competencies, social support, and interpersonal process in psychotherapy. *Psychotherapy Research, 10,* 239–266. http://dx.doi.org/10.1093/ptr/10.3.239

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 134–166). Hoboken, NJ: Wiley.

Overall, N. C., Fletcher, G. J., & Kenny, D. A. (2012). When bias and insecurity promote accuracy mean-level bias and tracking accuracy in couples' conflict discussions. *Personality and Social Psychology Bulletin, 38,* 642–655. http://dx.doi.org/10.1177/0146167211432764

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Rosenblatt, A., & Rosenblatt, J. A. (2002). Assessing the effectiveness of care for youth with severe emotional disturbances: Is there agreement between popular outcome measures? *Journal of Behavioral Health Services & Research, 29,* 259–273. http://dx.doi.org/10.1007/BF02287367

Sánchez-Meca, J., Rosa-Alcázar, A. I., Marín-Martínez, F., & Gómez-Conesa, A. (2010). Psychological treatment of panic disorder with or without agoraphobia: A meta-analysis. *Clinical Psychology Review, 30,* 37–50. http://dx.doi.org/10.1016/j.cpr.2009.08.011

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78,* 298–311. http://dx.doi.org/10.1037/a0019247

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. Newbury Park, CA: Sage.

Stuart, S., & Noyes, R., Jr. (1999). Attachment and interpersonal communication in somatization. *Psychosomatics: Journal of Consultation and Liaison Psychiatry, 40,* 34–43. http://dx.doi.org/10.1016/S0033-3182(99)71269-7

Subica, A. M., Fowler, J. C., Elhai, J. D., Frueh, B. C., Sharp, C., Kelly, E. L., & Allen, J. G. (2014). Factor structure and diagnostic validity of the Beck Depression Inventory-II with adult clinical inpatients: Comparison to a gold-standard diagnostic interview. *Psychological Assessment, 26,* 1106–1115. http://dx.doi.org/10.1037/a0036998

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74,* 242–261. http://dx.doi.org/10.1207/S15327752JPA7402_6

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome Questionnaire: Is it sensitive to changes in counselling center clients? *Journal of Counseling Psychology, 51,* 38–49. http://dx.doi.org/10.1037/0022-0167.51.1.38

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review, 118,* 357–378. http://dx.doi.org/10.1037/a0022936

Received January 7, 2016
Revision received August 9, 2016
Accepted August 26, 2016 ■