Online supplement

The following materials are intended as supplementary material. Due to space limitations, they are referenced from but not printed in the actual paper.

Method

Patients. Of the patients, 63.5% were employed (part-time or full time) and 71.5% were female. As to marital status, 88.2% of the patients were single, 9.6% were married, 1.7% were divorced, 0.2% were widowed, and 0.2% did not answer the question about their marital status. The country of origin of 83.1% is Israel; 9.6% were immigrants from the former Soviet Union, 2.5% were immigrants from Ethiopia, and 4.8% indicated "elsewhere" in the pre-treatment questionnaire.

Inclusion criteria at baseline were as follows: every applicant who gave his or her approval to use their personal data for research purpose was automatically included in the study. There is no quantitative data on the number of applicants who did not provide such approval, but the subjective impression of the clinic team was that the overwhelming majority of applicants provided the approval. The invitation to participate in the second and third stages of the study at the end of treatment did not depend on the length of treatment or dropout.

Professional evaluators. Evaluators who were interns also participated in an intensive training course, which focused mainly on a dynamic formulation and on conceptualization of the patient's self-understanding. Before receiving their first case, they attended weekly individual training sessions for several weeks, where they acquired the skill to evaluate the patient's clinical status (according to the DSM-IV-TR framework), including their insight level. After being assigned their first case, and in the course of their entire work of patient evaluation, they received weekly individual supervision sessions, where they discussed their rating decisions with their

supervisor. Professional evaluators who were licensed therapists had received this training when they were interns, and subsequently received weekly group supervision on their cases.

Measures

Insight based on professional evaluator's perspective.

The evaluation was performed twice: once at baseline, before the beginning of treatment, and once at the end of the treatment, by professional evaluators, independent psychologists of the counseling center (not the therapists), who performed interviews with the patients. In the initial assessments, inter-rater reliability was not assessed because of practical limitations: as is often the case in the clinical field, each intake interview was conducted only once, and was not audio-recorded. Hence, only a single evaluator could assess the patient's level of insight. The evaluations at the end of treatment were all performed by the same psychologist. The sessions were audio-recorded, and three psychologists listened to the recordings and rated the patients' level of insight. During the coding period, the psychologists listened together to the recordings of three sessions and discussed the ratings. There is no numeric quantification of their degree of reliability, but their subjective assessment at the time of the meetings was that the degree of agreement was very high.

The Hebrew wording of the term "self-understanding" presented to the professional evaluator was slightly different from the Hebrew wording of this term presented to the patient. This difference in the wording was chosen to match their different backgrounds (education, experience, etc.). For therapists, we used a professional term, whereas for patients, we used non-professional language. Both patients and evaluators were asked to assess the extent to which patients were aware of their inner world and understood the reasons for their various behaviors. This choice is not uncommon, and has been made in previous studies that focused on patients' and therapist's agreement and disagreement (Chui et al., 2019).

Data analysis

The statistical method was succinctly described in the article itself, and is fully described below.

To examine our hypotheses, we conducted a response surface analysis by polynomial regression (Shanock et al., 2010). This statistical method was introduced to the field of counseling and psychotherapy research by Marmarosh and Kivlighan (2012), and since then has been used in more than a dozen psychotherapy research studies.

When examining the initial level of insight (the first and the second research questions), we used polynomial regression equations that consisted of five predictors: (a) patient's initial insight rating, (b) professional evaluator's initial insight rating, (c) a quadratic term formed by squaring the patient's insight rating, (d) a quadratic term formed by squaring the professional evaluator's insight rating, and (e) a cross-product term formed by multiplying the patient insight rating with the professional evaluator insight rating. Both patient's and professional evaluator's insight ratings were centered around their midpoint. The following regression equations were used:

Research question 1- ***Baseline insight and pre-treatment symptoms***:

$$SY = b_0 + b_1PA + b_2PE + b_3PA^2 + b_4PA \times PE + b_5\,PE^2 + e,$$

Research question 2- ***Baseline insight and changes in symptoms***:

$$\Delta SY = b_0 + b_1PA + b_2PE + b_3PA^2 + b_4PA \times PE + b_5\,PE^2 + e,$$

where PA represents the patient's insight rating, PE the professional evaluator's insight rating, SY the patient's symptoms, and Δ the change in the variable from the beginning of treatment to its termination point.

When examining the change in insight during treatment (the third research question), the polynomial regression equation consisted of five similar predictors, but in this case we used the change in insight during treatment, rather than the initial level of insight. Moreover, two variables were added to the equation to control for the initial level of insight. Patients rated as having a high level of insight at the beginning of treatment had little chance of improving insight during the course of treatment and, therefore, could have biased the results of this regression with a ceiling effect. Therefore, we carried out this regression in steps. In the first step, the following two variables were added: (a) patient's initial insight rating and (b) professional evaluator's initial insight rating. In the second step, these two variables were added again, together with the five variables of the surface analysis, as described above. Both patient's and professional evaluator's changes in insight ratings were centered around their midpoint. The following regression equation was used:

Research question 3. ***Changes in insight and changes in symptoms:***

Step 1:

$$\Delta SY = b_0 + b_1 PA + b_2 PE + e,$$

Step 2:

$$\Delta SY = b_0 + b_1 PA + b_2 PE + b_3(\Delta PA) + b_4(\Delta PE) + b_5(\Delta PA)^2 + b_6(\Delta PA) \times \Delta PE + b_7 (\Delta PE)^2 + e,$$

where PA represents the patient's insight rating, PE the professional evaluator's insight rating, SY the patient's symptoms, and Δ the change in the variable from the beginning of treatment to its termination point.

We used the estimated coefficients from the regression model to calculate test values for two slopes and two curvatures along the response surface (for the sake of space, the examples within brackets relate to the first research question only): (a) the slope of the line of agreement

(which tests whether symptoms are less severe when patients and professional evaluators agree that the insight level is high than when they agree that the insight level is low; $a_1 = b_1 + b_2$); (b) the curvature along the line of agreement (which tests whether symptoms are less severe when patients and professional evaluators agree that the insight level is moderate than when they agree that the insight level is ether high or low; $a_2 = b_3 + b_4 + b_5$); (c) the slope of the line of disagreement (which tests whether symptoms are less severe when the patient's perceptions of insight are higher than the professional evaluator's perceptions of insight, than when the professional evaluator's perceptions of insight are higher than the patient's perceptions; $a_3 = b_1 - b_2$); and (d) the curvature along the line of disagreement (which tests whether symptoms are less severe when the degree of disagreement between patient and professional evaluator ratings of insight is small; $a_4 = b_3 - b_4 + b_5$). For more information, see Shanock, Baran, Gentry, Pattison, and Heggestad (2010).

Results

Preliminary analysis

A table listing the correlations between all variables used in the main analyses, including the level of insight, the level of symptoms and the length of treatment, is presented in Appendix A (Table A1). The table also presents the average and standard deviation of the various variables.

To determine whether there was an association between the demographic characteristics of the patients and their choice to cooperate with the assessments made at the end of the treatment, Chi-square tests for association were conducted. We divided the sample into three subgroups: (a) patients who had an intake assessment only, (b) patients who answered the symptom questionnaire at the end of the treatment, and (c) patients who had an evaluation session with the evaluator at the end of treatment. We used the Chi-square test for association to determine if there was an association between the subgroup to which the patient belongs and his or her different demographic characteristics. The comparison showed no association between the

subgroup to which the subject belongs and most of the demographic characteristics assessed. No association was found with gender, education, marital status, or employment. One characteristic of the participants was found to be related to the sample to which they belonged: belonging to a minority group χ^2 (8, 416) = 17.49, p = 0.03. This finding may indicate that minorities were less likely to cooperate in the follow-up segment of the study. We addressed this in the discussion as a limitation.

To compere the effect of patient's choice to cooperate with the assessments made at the end of the treatment on self-understanding variables at bassline, a one-way ANOVA test was conducted. The results of these tests were not significant for both the patient's perspective regarding his level of insight at baseline ($F_{(2,410)} = 0.56$, $p = 0.57$), and for the professional evaluator perspective regarding the patient's level of insight at baseline ($F_{(2,394)} = 0.08$, $p = 0.93$).

Following Shanock et al. (2010), we tested whether the discrepancy between the patient's and the professional evaluator's insight ratings was high enough (more than about 50%). For illustration purposes, if we had found that the vast majority of patients fully agreed with the professional evaluators regarding their level of insight (a situation of low discrepancy), there would have been no point in conducting surface analysis. The findings show that 68% of the patients had discrepancy values for the sample of Hypothesis 1, 57% for the sample of Hypothesis 2, and 70% for the sample of Hypothesis 3. Therefore, we proceeded with the surface analyses to test each of the three hypotheses.

Missing data management

The database of the counseling center consists of records of 417 patients. At baseline, two patients (0.5%) did not fill the ComPASS questionnaire and four patients (1.0%) did not answer the insight question. Furthermore, at baseline, 20 evaluators (4.8%) did not answer the insight question. At baseline, data was missing completely at random (Little's MCAR test $X^2 = 4.01$, df

= 6, $p = 0.68$, n.s). At the termination, one patient did not fill the ComPASS questionnaire (1%) and the insight question (1%). At termination, data was missing completely at random (Little's MCAR test $X^2 = 1.472$, df = 1, $p = 0.26$. n.s)

To address the potential impact of missing data, we repeated the main statistical analyses on a data set where missing data were imputed. We used the multiple imputation procedure in the SPSS statistical package, version 26, to impute the missing data. Missing data were found in all the variables we used in all three research questions (the levels of insight and symptoms before and after treatment, from both the patients' and the evaluators' perspective). All variables were included in the imputation model, and were also used as predictors in this model. The maximum number of draws per case was specified at 50 and maximum parameter draws at 2. Five imputed datasets were created.

Following the imputation, we repeated the main analyses reported in the Results section on the five new data sets. The results showed a level of significance highly similar to that of the original analysis. Therefore, we decided to report the original results, rather than the results of the analysis of the imputed datasets.

References

Chui, H., Palma, B., Jackson, J. L., & Hill, C. E. (2019). Therapist–client agreement on helpful and wished-for experiences in psychotherapy: Associations with outcome. *Journal of Counseling Psychology*.

Marmarosh, C. L., & Kivlighan, D. M. (2012). Relationships among client and counselor agreement about the working alliance, session evaluations, and change in client symptoms using response surface analysis. *Journal of Counseling Psychology, 59*(3), 352.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*(4), 543–554.

Appendix A

Table A1

Means, standard deviations and correlations for insight, symptoms and treatment length

	M	SD	1	2a	2b	2c	3a	3b	3c	4a	4b
1. Length of treatment (weeks)	36.85	28.37									
2. Insight level, patient's perspective											
2a. Pre-treatment	2.66	1.00	-.11								
2b. post-treatment	3.84	1.01	-.14	.20							
2c. Change during treatment	0.25	1.20	.02	-.57**	.70**						
3. Insight level, professional evaluator perspective											
3a. Pre-treatment	1.82	0.76	.09	.10	.11	-.05					
3b. post-treatment	3.10	0.78	.18	.03	.09	.04	.56**				
3c. Change during treatment	0.26	0.76	.09	-.34	-.45**	-.13	-.53**	.44*			
4. Psychopathological Symptoms											
4a. Pre-treatment	2.65	0.88	.05	-.35**	-.260*	.01	-.14**	.17	.32		
4b. post-treatment	1.86	0.62	-.04	-.06	-.32**	-.21*	-.05	-.16	-.01	.47**	
4c. Change during treatment	0.73	0.83	.08	-.26*	-.04	.17	-.07	.27	.31	.76**	-.22*

Note. M=Mean, SD= standard deviation.

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A2

Polynomial Regression Model for baseline insight from the perspectives of the patient and the professional evaluator, and patient's pre-treatment symptoms.

Effect	Estimate	S.E.	*t*(387)	*p*
Symptom severity intercept	2.57	0.08	33.6	0.00
Insight**,** patient's perspective	-0.40	0.06	-6.66	0.00
Insight, evaluator's perspective	-0.18	0.08	-2.36	0.02
Insight, patient's perspective [a]	0.17	0.04	4.46	0.00
Insight, patient's × evaluator's perspectives	0.04	0.07	0.61	0.54
Insight, evaluator's perspective [b]	0.18	0.06	2.79	0.01

Note. SE= Standard error.

[a] Quadratic term formed by squaring the patient's insight rating.

[b] Quadratic term formed by squaring the evaluator's insight rating.

Table A3

Response surfaces for patient's and professional evaluator's perspective on the baseline level of insight and the change in insight, and patient's pre-treatment symptoms and change in symptoms.

Research question	Effect	B	S.E.	t[a]	p
Baseline insight and pre-treatment symptoms	a_1: Slope along x = y	-0.58	0.07	-8.20	0.00
	a_2: Curvature along x = y	0.35	0.07	4.72	0.00
	a_3: Slope along x = -y	-0.22	0.10	-2.34	0.02
	a_4: Curvature along x = -y	0.17	0.10	1.73	0.09
Baseline insight and changes in symptoms	a_1: Slope along x = y	-0.49	0.15	-3.16	0.00
	a_2: Curvature along x = y	0.29	0.14	2.02	0.05
	a_3: Slope along x = -y	-0.37	0.22	-1.69	0.10
	a_4: Curvature along x = -y	0.22	0.22	1.02	0.31
Changes in insight and changes in symptoms	a_1: Slope along x = y	-0.26	0.42	-0.62	0.54
	a_2: Curvature along x = y	1.18	0.46	2.56	0.01
	a_3: Slope along x = -y	-0.26	0.41	-0.64	0.52
	a_4: Curvature along x = -y	0.11	0.31	0.35	0.73

Note. S.E.= Standard error.

[a] The degrees of freedom were 387 for the baseline insight and pre-treatment symptoms, 88 for the baseline insight and changes in symptoms, and 25 for the changes in insight and changes in symptoms.

Table A4

Polynomial Regression Model for the baseline insight from the perspectives of the patient and the professional evaluator, and patient's change in symptoms.

Effect	Estimate	S.E.	*t*(88)	*p*
Change in symptoms intercept	0.72	0.13	5.58	0.00
Insight**,** patient's perspective	-0.43	0.13	-3.23	0.00
Insight, evaluator's perspective	-0.06	0.14	-0.44	0.66
Insight, patient's perspective [a]	0.20	0.09	2.35	0.02
Insight, patient's × evaluator's perspectives	0.03	0.12	0.25	0.80
Insight, evaluator's perspective [b]	0.05	0.11	0.51	0.62

Note. SE= Standard error.

[a] Quadratic term formed by squaring the patient's insight rating.

[b] Quadratic term formed by squaring the evaluator's insight rating.

Table A5

Polynomial Regression Model for the baseline level of insight from the perspectives of the patient and the professional evaluator, and patient's change in symptoms.

Model	Effect	Estimate	S.E.	*t*(25)	*p*
1	Change in symptoms intercept	0.71	0.21	3.42	0.00
	Insight, patient's perspective	-0.05	0.20	-0.24	0.81
	Insight, evaluator's perspective	-0.04	0.20	-0.19	0.85
2	Change in symptoms intercept	0.32	0.31	1.03	0.31
	Change in insight, patient's perspective	-0.26	0.24	-1.11	0.28
	Change in insight, evaluator's perspective	0.00	0.34	0.00	0.10
	Change in insight, patient's perspective [a]	0.20	0.12	1.63	0.12
	Change in insight, patient's * evaluator's perspectives	0.54	0.26	2.09	0.05
	Change in insight, evaluator's perspective [b]	0.44	0.26	1.69	0.10

Note. SE= Standard error.

[a] Quadratic term formed by squaring the patient's insight rating.

[b] Quadratic term formed by squaring the evaluator's insight rating.

Appendix B

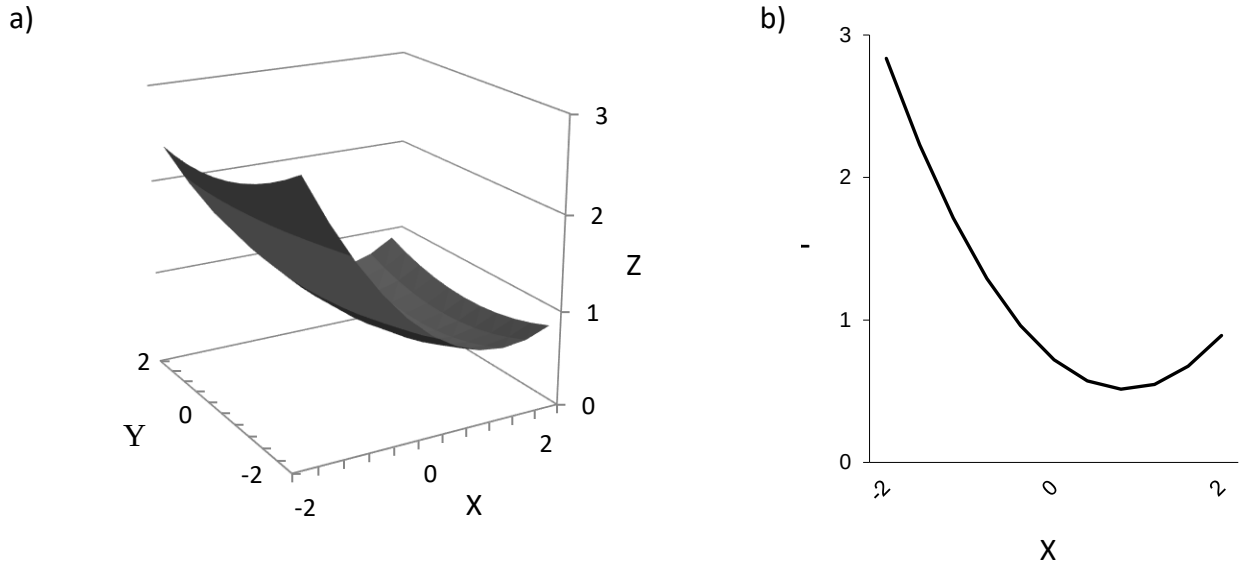

Figure B1. Patient's and professional evaluator's perspective on the patient's initial level of insight, and the change in symptoms. In (a) the X-axis is the patient's perspective on his level of initial insight (centered), the Y-axis is the professional evaluator's perspective on the patient's initial level of insight (centered), and the Z-axis is the change in symptoms (positive numbers represent an improvement in the patient's feeling). In (b) the x-axis is the patient's insight (given agreement between patient's and professional evaluator's insight ratings) and the y-axis is the change in symptoms (positive numbers represent an improvement in the patient's feeling).

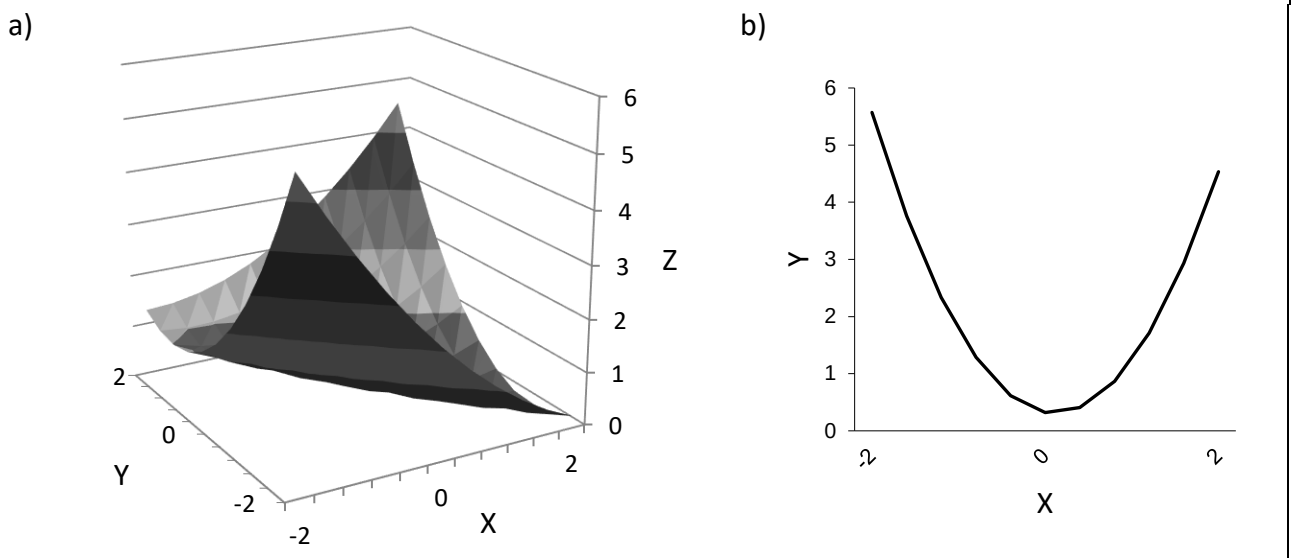

Figure B2. The change in patient's and professional evaluator's perspectives regarding the patient's level of insight and the change of symptoms. In (a) the X-axis is the change in patient's perspective on his own level of insight (centered), the Y-axis is the change in the professional evaluator's perspective on the patient's level of insight (centered), and the Z-axis is the change in symptoms (positive numbers represent an improvement in the patient's feeling). In (b) the x-axis is the change in patient's insight (given agreement between patient's and professional evaluator's insight ratings) and the y-axis is the change in symptoms (positive numbers represent an improvement in the patient's feeling).