**Online Supplemental Information**

**Table S1**

*A Summary of the ICC for the Individual Rating Groups with and without the Primary Rater Included.*

|  |  |  |
| --- | --- | --- |
|  | Rating Group A (6 recordings) | Rating Group B (6 recordings) |
|  | Rating Group Alone | Rating Group and Primary Rater | Rating Group Alone | Rating Group and Primary Rater |
| Overall Score only (95%CI) | 0.92 (0.68 – 0.99) | 0.85 (0.53 – 0.98) | 0.86 (0.34 – 0.98) | 0.90 (0.6 – 0.98) |
| Items and Overall Score (95%CI) | 0.98 (0.97 – 0.99) | 0.98 (0.97 – 0.99) | 0.99 (0.98 – 0.99) | 0.99 (0.99 – 1.0) |

*Notes*. ‘Absolute agreement’ was used for ICC. CI = Confidence Intervals.

**Table S2**

*Summaries of Outcome Change at End of Treatment for the Overall Sample and Each Treatment Condition.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Overall (n=94) | CBT-GSH (n=23) | CAT-GSH (n=71) | Test of Difference |
| Anxiety (GAD-7) |  |  |  |  |
| Reliable improvement | 66(70.2%) | 19 (82.6%) | 47(66.2%) | χ2 (1) = 2.24, *p =* 0.135 |
| Reliable deterioration | 1 (1.1%) | 0 | 1 (1.4%) | χ2 (1) = 0.33, *p =* 0.567 |
| RCSI | 51 (54.3%) | 16 (69.6%) | 35 (49.3%) | χ2 (1) = 2.88, *p =* 0.09 |
| No change | 30 (31.9%) | 5 (21.7%) | 25 (35.2%) | χ2 (1) = 1.45, *p =* 0.228 |
| Depression (PHQ-9) |  |  |  |  |
| Reliable improvement | 49 (52.1%) | 17(73.9%) | 32 (45.1%) | χ2 (1) = 5.79, *p =* 0.016\* |
| Reliable deterioration | 3 (3.2%) | 0 | 3 (4.2%) | χ2 (1) = 1, *p =* 0.316 |
| RCSI | 35 (37.2%) | 13(56.5%) | 22 (31%) | χ2 (1) = 4.85, *p =* 0.028\* |
| No change | 39 (41.5%) | 5 (21.7%) | 34 (47.9%) | χ2 (1) = 4.89, *p =* 0.027\* |

*Notes.* \*significant at p < .05 threshold, \*\*significant at p < .01 threshold, \*\*\*significant at p < .001 threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

**Table S3**

*Chi-Square Analyses for Rates of RCSI Across PWP Competence Levels per Treatment Condition and for the Overall Sample.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Overall Sample (*n*=94) | CAT-GSH Condition (*n*=71) | CBT-GSH Condition (*n*=23) |
|  |  | Low1  | Medium2 | High3 | Low | Medium | High | Low | Medium | High |
|  | Total N | 24 | 48 | 22 | 20 | 37 | 14 | 4 | 11 | 8 |
| Anxiety |  |
|  | RCSI | 15 | 24 | 12 | 11 | 17 | 7 | 4 | 7 | 5 |
|  | No RCSI | 9 | 24 | 10 | 9 | 20 | 7 | 0 | 4 | 3 |
|  | Chi-square | χ2 (2) = 1.01, p = 0.604 | χ2 (2) = 0.43, *p* = 0.807 | χ2 (2) = 2.12, *p* = 0.346 |
| Depression |  |
|  | RCSI | 10 | 17 | 8 | 7 | 11 | 4 | 3 | 6 | 4 |
|  | No RCSI | 14 | 31 | 14 | 13 | 26 | 10 | 1 | 5 | 4 |
|  | Chi-square | χ2 (2) = 0.28, p = 0.871 | χ2 (2) = 0.22, *p* = 0.9 | χ2 (2) = 0.71, *p* = 0.701 |

*Notes.* RCSI = Reliable and Clinically Significant Improvement. Significance is Asymptotic Significance (two-sided). 1

= low competence was defined as an overall competence score < 17.5, 2 = medium competence was defined as an overall score between 17.5 and 21.5, 3 = high competence was defined as an overall score >21.5. \*significant at p < .05 threshold, \*\*significant at p < .01 threshold, \*\*\*significant at p < .001 threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

**Table S4**

*Hierarchical Regression Model for Association Between Competence and Anxiety Outcome, Controlling for Baseline Severity, Treatment Condition and Timing of Session Rating.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *B* | *B* SE | $$β$$ | *p* |
| Step 1 (*R2*=0.13) |  |  |  |  |
|  | *Constant* | 3.31 | 3.25 | - | 0.311 |
|  | Baseline GAD-7 | 0.33\* | 0.13 | 0.24 | 0.016 |
|  | Treatment condition | 1.31 | 1.22 | 0.11 | 0.283 |
|  | Session number rated | -0.8\* | 0.31 | -0.25 | 0.012 |
| Step 2 (*R2*=0.15) |  |  |  |  |
|  | *Constant* | -1.81 | 4.83 | - | 0.710 |
|  | Baseline GAD-7 | 0.29\* | 0.14 | 0.21 | 0.035 |
|  | Treatment condition | 1.6 | 1.23 | 0.13 | 0.196 |
|  | Session number rated | -0.8\* | 0.31 | -0.25 | 0.011 |
|  | Competence | 0.27 | 0.19 | 0.14 | 0.158 |
| Step 3 (*R2*=0.15) |  |  |  |  |
|  | *Constant* | -11.76 | 16.87 | - | 0.488 |
|  | Baseline GAD-7 | 0.29 | 0.14 | 0.22 | 0.035 |
|  | Treatment condition | 7.16 | 9.11 | 0.58 | 0.434 |
|  | Session number rated | -0.81\* | 0.31 | -0.26 | 0.011 |
|  | Competence | 0.77 | 0.83 | 0.41 | 0.358 |
|  | Competence\*Treatment | -0.28 | 0.45 | -0.49 | 0.540 |

*Notes.* Step 2: Δ*R2* = 0.01. Step 3Δ*R2* = 0.001. \* *p*<.05, \*\* *p*<.01, \*\*\**p*<.001. SE = Standard Error.

**Table S5**

*Sensitivity Analysis for Hierarchical Regression Model for Association Between Competence and Anxiety Outcome, Controlling for Baseline Severity, Treatment Condition and Timing of Session Rating (only session number rated 1 or 2 included)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *B* | *B* SE | $$β$$ | *p* |
| Step 1 (*R2*=0.08) |  |  |  |  |
|  | *Constant* | 2.32 | 6.18 | - | 0.709 |
|  | Baseline GAD-7 | 0.36 | 0.21 | 0.27 | 0.096 |
|  | Treatment condition | 0.93 | 1.88 | 0.07 | 0.625 |
|  | Session number rated | -0.12 | 1.73 | -0.01 | 0.946 |
| Step 2 (*R2*=0.10) |  |  |  |  |
|  | *Constant* | -5.06 | 9.21 | - | 0.586 |
|  | Baseline GAD-7 | 0.34 | 0.21 | 0.25 | 0.113 |
|  | Treatment condition | 1.35 | 1.92 | 0.11 | 0.486 |
|  | Session number rated | 0.25 | 1.76 | 0.02 | 0.888 |
|  | Competence | 0.33 | 0.31 | 0.17 | 0.286 |
| Step 3 (*R2*=0.10) |  |  |  |  |
|  | *Constant* | -1.36 | 25.54 | - | 0.958 |
|  | Baseline GAD-7 | 0.34 | 0.21 | 0.25 | 0.12 |
|  | Treatment condition | -0.85 | 14.24 | -0.07 | 0.953 |
|  | Session number rated | 0.31 | 1.83 | 0.03 | 0.865 |
|  | Competence | 0.14 | 1.28 | 0.07 | 0.914 |
|  | Competence\*Treatment | 0.11 | 0.72 | 0.19 | 0.877 |

*Notes.* Step 2: Δ*R2* = 0.01. Step 3Δ*R2* = 0.001. \* *p*<.05, \*\* *p*<.01, \*\*\**p*<.001. SE = Standard Error.

**Table S6**

*Chi-Square Analyses for Session Attendance Across PWP Overall Competence Levels per Treatment Condition and for the Overall Sample.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Overall Sample (*n*=94) | CAT-GSH Condition (*n*=71) | CBT-GSH Condition (*n*=23) |
| Attendance |  | Low | Medium | High | Low | Medium | High | Low | Medium | High |
|  | Total N | 24 | 48 | 22 | 20 | 37 | 14 | 4 | 11 | 8 |
|  | Minimal attendance | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 |
|  | Moderate attendance | 4 | 8 | 5 | 3 | 5 | 2 | 1 | 3 | 3 |
|  | Full attendance | 20 | 37 | 17 | 17 | 29 | 12 | 3 | 8 | 5 |
|  | Chi-square | χ2 (4) = 3.3, *p* = 0.508 | χ2 (4) = 2.88, *p* = 0.578 | χ2 (2) = 0.3, *p* = 0.862 |

*Notes.* \*significant at p < .05 threshold, \*\*significant at p < .01 threshold, \*\*\*significant at p < .001 threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

**Table S7**

*Binomial Logistic Regression Model for Association Between Competence and Drop-Out, Controlling for Baseline Severity, Treatment Condition and Timing of Session Rating.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *B* | *B* SE | Exp(B) | *p* |
| Step 1 (*Nagelkerke R2=0.23*) |  |  |  |  |
|  | *Constant* | -3.3 | 1.95 | 0.04 | 0.090 |
|  | Baseline GAD-7 | 0.17 | 0.1 | 1.19 | 0.097 |
|  | Session number rated | -0.69\* | 0.3 | 0.5 | 0.019 |
| Step 2 (*Nagelkerke R2=0.24*) |  |  |  |  |
|  | *Constant* | -3.23 | 3.5 | 0.04 | 0.357 |
|  | Baseline GAD-7 | 0.18 | 0.11 | 1.2 | 0.084 |
|  | Session number rated | -0.7\* | 0.3 | 0.5 | 0.018 |
|  | Treatment condition | 0.51 | 0.88 | 1.66 | 0.562 |
|  | Competence | -0.06 | 0.12 | 0.95 | 0.645 |
| Step 3 (*Nagelkerke R2=0.28*) |  |  |  |  |
|  | *Constant* | -20.78 | 16.67 | 0.00 | 0.213 |
|  | Baseline GAD-7 | 0.17 | 0.1 | 1.18 | 0.098 |
|  | Treatment condition | 10.21 | 8.73 | 27074 | 0.242 |
|  | Session number rated | -0.74\* | 0.31 | 0.48 | 0.017 |
|  | Competence | 0.81 | 0.78 | 2.24 | 0.299 |
|  | Competence\*Treatment | -0.47 | 0.41 | 0.62 | 0.246 |

*Notes.* Step 2: *Nagelkerke R2* = 0.06. Step 3 *Nagelkerke R2* = 0.01. \* *p*<.05, \*\* *p*<.01, \*\*\**p*<.001. SE = Standard Error.

**Table S8**

*Sensitivity Analysis for Binomial Logistic Regression Model for Association Between Competence and Drop-Out, Controlling for Baseline Severity, Treatment Condition and Timing of Session Rating (only session number rated 1 or 2 included).*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *B* | *B* SE | Exp(B) | *p* |
| Step 1 (*Nagelkerke R2=0.07*) |  |  |  |  |
|  | *Constant* | -2.1 | 2.36 | 0.12 | 0.373 |
|  | Baseline GAD-7 | 0.1 | 0.11 | 1.1 | 0.355 |
|  | Session number rated | -0.65 | 0.81 | 0.52 | 0.421 |
| Step 2 (*Nagelkerke R2=0.08*) |  |  |  |  |
|  | *Constant* | -2.08 | 4.26 | 0.13 | 0.625 |
|  | Baseline GAD-7 | 0.1 | 0.11 | 1.11 | 0.346 |
|  | Session number rated | -0.65 | 0.83 | 0.52 | 0.435 |
|  | Treatment condition | 0.2 | 0.92 | 1.22 | 0.826 |
|  | Competence | -0.02 | 0.14 | 0.98 | 0.873 |
| Step 3 (*Nagelkerke R2=0.14)* |  |  |  |  |
|  | *Constant* | -20.33 | 17.38 | 0.00 | 0.242 |
|  | Baseline GAD-7 | 0.08 | 0.11 | 1.09 | 0.436 |
|  | Treatment condition | 10.79 | 9.36 | 48279.27 | 0.249 |
|  | Session number rated | -0.96 | 0.88 | 0.38 | 0.274 |
|  | Competence | 0.9 | 0.83 | 2.47 | 0.274 |
|  | Competence\*Treatment | -0.52 | 0.44 | 0.6 | 0.238 |

*Notes.* Step 2: *Nagelkerke R2* = 0.06. Step 3 *Nagelkerke R2* = 0.01. \* *p*<.05, \*\* *p*<.01, \*\*\**p*<.001. SE = Standard Error.

**Table S9**

*Chi-Square Analyses for Stepping up Rate Across PWP Competence Levels per Treatment Condition and for the Overall Sample.*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Overall Sample (*n*=94) | CAT-GSH Condition (*n*=71) | CBT-GSH Condition (*n*=23) |
|  | Low | Medium | High | Low | Medium | High | Low | Medium | High |
| Total N | 24 | 48 | 22 | 20 | 37 | 14 | 4 | 11 | 8 |
| Stepped up | 5 | 6 | 4 | 4 | 6 | 4 | 1 | 0 | 0 |
| Not stepped up | 19 | 42 | 18 | 16 | 31 | 10 | 3 | 11 | 8 |
| Chi-square | χ2 (2) = 0.93, *p* = 0.627 | χ2 (2) = 0.98, *p* = 0.612 | χ2 (2) = 4.97, *p* = 0.083 |

*Notes.* \*significant at p < .05 threshold, \*\*significant at p < .01 threshold, \*\*\*significant at p < .001 threshold, significant at Bonferroni adjusted threshold for multiple testing in bold (though none are significant at the adjusted threshold).

**Sensitivity Analyses from Regressions**

**Sensitivity Analysis for Multiple Regression**

In the multiple regression model, the ‘session number rated’ covariate had a significant main effect on end of treatment GAD-7 score when all other variables in the model were controlled for, B = -0.76 (95% CI = -1.45 – -0.06), SE = 0.35, *p* = 0.03. To help elucidate whether PWP competence influenced clinical outcome, or clinical outcome influenced PWP competence, a sensitivity analysis was conducted with only sessions that were rated early in treatment (rated at session 1 or 2) being included in the multiple regression model (n=39). The main effect of overall competence on end of treatment GAD-7 score remained non-significant with treatment, baseline GAD-7 score and session number rated being controlled for, B = 0.16 (95% CI = -0.63 – 0.95), SE = 0.39, *p =* 0.68.

**Sensitivity Analysis for Logistic Regression**

A sensitivity analysis was also conducted for the logistic regression, where the same early-rated sample were included in the logistic regression model (n=39). The main effect of overall competence on patient drop-out remained non-significant with baseline GAD-7 score, session number rated, treatment condition and competence by treatment interaction being controlled for, B (1) = -0.27, SE = 0.21, *p* = 0.20.

**LMM Model-Building Procedure**

**Unconditional Models**

***Individual Growth Trajectories***

After an unconditional model with a simple variance components structure was run, a number of different individual growth trajectories were tested to establish the best fitting time trend. After testing linear, log-linear, quadratic and cubic models (all with time as fixed effects); the log-linear model was selected. Chi-square analyses were used to establish whether the -2 Log-Likelihood were significantly different. Consistent with the rule of parsimony, the simplest model with the best fit statistics was selected. The log-linear model was the simplest (4 parameters) model with the best fitting -2 Log-Likelihood statistic which was less than the linear model statistic. A chi-squared test was unnecessary due to the log-linear model having the same number of parameters as the linear model. The more complex structures (i.e. quadratic and cubic) did not fit the data significantly better than log-linear and so the log-linear structure was taken forward.

After an unconditional fixed growth model with log-linear trend was run, an unconditional model with log-linear (time trend) as random effects was run. This was in order to test whether allowing intercepts and slopes to vary would improve the model. This resulted in substantial improvement in model fit which explained an additional 13% of the residual variance.

***Covariance Structure***

Next, the best-fitting within-individual error covariance structure was selected before entering any predictors. The best fitting covariance structure was a AR1: Heterogenous structure (modelling for auto-correlation in the longitudinal data).

**Conditional Models (Adding Covariates and Predictor)**

After unconditional models were run, the conditional models were run with overall competence (continuous) and time (continuous) entered into the LMM analyses as predictors and session number rated and treatment group entered as covariates. Before either covariates or the predictor were entered, they were mean-centred, as is recommended in LMM (Shek & Ma, 2011).

The covariates were entered first in order to see their individual main effects. Adding session number rated (fixed effect) significantly improved the model fit (-2LL was significantly reduced), however no additional residual variance was explained. Adding treatment (fixed effect) to the model did not significantly improve model fit, nor did it explain any additional residual variance.

Adding overall competence as a fixed effects predictor significantly improved the model fit. This model only explained an additional 0.1% of residual variance. A patient-level interaction term (competence \* time) was then added to the model in order to test whether the between-patient variation in change in GAD-7 scores over time was explained by the competence of their PWP, with session number rated and treatment condition being controlled for. The model fit was not significantly improved compared to the model with only covariates included. The level of PWP competence, therefore, did not appear to have a significant impact on the rate of change in self-reported anxiety of patients through GSH intervention for the overall sample.

**Table S10**

*Statistics at each stage of the LMM model-building process:*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Deviance (-2LL), df** | **Change in Deviance, df**  | **Residual variance****(% additional variance explained)** | **Subject level intercept variance**  | **Subject level slope variance for time** | **Intercept-Slope Covariances** |
| 1. Unconditional model | 3832, 3 | ---  | 17.03 | 11.81 | --- | --- |
| 2. Unconditional fixed linear growth model (time only as fixed effect) – VC  | 3590, 4 | 242\*, 1  | 11.05 (35.11%) | 12.40 | --- | --- |
| 3. Unconditional fixed growth model with high-order trends of time |  |  |  |  |  |  |
| **Log-linear**  | **3575, 4** | **15, 0**  | **10.77****(0.02%)** | **12.31** | **---** | **---** |
| Quadratic  | 3575, 5 | 0, 1  | 10.77(0%) | 12.29 | --- | --- |
| Cubic  | 3573, 6 | 2, 2 | 10.72 (0%) | 12.36 | --- | --- |
| **4. Unconditional random growth model (best fitting time trend as random effects) (VC)** | **3532, 5** | **42, 0** | **8.69** **(12%)** | **10.42** | **3.69** | **---** |
| Testing different covariance structures |  |  |  |  |  |  |
| **AR1 Heterogenous** | **3525, 6** | **7\*, 1** | **8.36** **(2%)** | **13.86** | **5.12** | **-0.40** |
| Unstructured | 3525, 6 | 0, 0 | 8.36(0%) | 13.86 | 5.12 | -3.39  |
| 5. Adding main effects of covariates (session rated and treatment) |  |  |  |  |  |  |
| Session number rated as covariate | 3505, 7 | 20\*, 1  | 8.40 (-0.2%) | 14.12 | 5.16 | -0.42 |
| Treatment group as covariate | 3505, 8 | 0, 1  | 8.39(0.06%) | 13.90 | 5.16 | -0.41 |
| 6. Adding main effect of competence | 3500, 9 | 5\*, 1  | 8.39(0%) | 12.87 | 5.17 | -0.41 |
| 7. Adding competence\*time interaction | 3499, 10 | 1, 1 | 8.40(-0.06%) | 12.84 | 5.12 | -0.41 |

*Notes.* \* = chi-square is significant at *p* < 0.05 level. The model taken forward to the subsequent phases is in bold. VC = variance components (covariance structure).