The evaluation of supervision: Construction of brief questionnaires for the supervisor and the supervisee

GERHARD ZARBOCK1, MARKO DREWS1, ALEXANDER BODANSKY1, & BERNHARD DAHME2

1Institut für Verhaltenstherapie-Ausbildung Hamburg, Hamburg & 2University of Hamburg, Hamburg, Germany

(Received 13 September 2007; revised 28 October 2008; accepted 9 December 2008)

Abstract
This study evaluated the psychometric properties of two parallel questionnaires (for the supervisor and supervisee, respectively). Ninety supervisees and 37 supervisors from different parts of Germany took part in the study. The three dimensions of clarifying, problem solving, and relationship were confirmed, but for both supervisor and supervisee versions of the scale medium intercorrelations also allowed a combined scale to be formed. In exploratory regression analysis, the relationship dimension related to both perspectives (i.e., supervisor and supervisee) served as the best predictor for overall supervision satisfaction. Despite the fact that general session satisfaction correlated in the medium range between supervisor and supervisee, there was no correlation between the perspectives as measured by the questionnaires. Possible reasons for this unexpected correlational pattern are discussed.

Keywords: supervision process; psychometric evaluation; supervision questionnaire

Supervision is used as a tool in three different settings. First, supervision is an essential part of the psychotherapy training curriculum designed to acquire the mandatory therapeutic skills (Frank, 1999). Second, for licensed therapists, continuous supervision is considered important to ensure quality management and professional integrity. Third, when training not yet fully licensed therapists, supervision helps to ensure that therapy is done lege artis and may help the supervisee cope with treatment problems and potential work crises. Single-case supervision can fulfill these different functions by supplying an external perspective on the process of psychotherapy and the factors influencing it (e.g., type of disorder, somatic comorbidity, treatment setting, competence and adherence of the therapist, and specific treatment method).

Within supervision, a number of different goals and functions can be distinguished, including those that are instructive (training), supportive (burn-out prophylaxis), and administrative (quality management; Schmelzer, 1997, p. 37). In general, supervision takes place in a dialectic range between support and control (Zimmer, 2000). Different procedures, goals, and methods of supervision can be applied in accordance with the level of competence of the supervisee and his or her individual needs. The needs of therapists concerning supervision depend on the current developmental phase of their individual training process and the aspects and characteristics that are predominant in that phase (Frank, Eucker, & Vaitl, 1988).

In an empirical analysis, Frank and Vaitl (1986) identified five needs of the supervisee: (1) need for emotional support, (2) need to enhance the supervisee’s own therapeutic competence, (3) need for an open emotional environment to facilitate self-exploration, (4) need to get case- and method-related guidance, and (5) need for activation and intellectual challenge.

Given the practical importance of supervision, the empirical research for its effectiveness is unsatisfactory (Auckenthaler, 1999; Ellis, Ladany, Krengel, & Schult, 1996; Falender & Shafranske, 2006; Schigl, 2004; Schmelzer, 1997). The question whether supervision has positive effects on supervised therapies has rarely been investigated; when it has been done, ambiguous results were reported (see overview by Bambling, King, Raue, Schweitzer, & Lambert, 2006; Lambert & Ogles, 1997; Frank, 1999; Sipos, 2001). However, supervision does seem to have some positive effect. Willutzki, Tönnies, and Meyer (2005)
studied the impact of 65 supervision sessions on the therapeutic bond. Sixteen therapists with 104 patients and six supervisors were included in this study. The treated patients suffered from anxiety and depressive disorders. The therapists were graduate trainees at a university center for psychological psychotherapy. The main finding of this study was a difference between how the therapeutic bond was rated in the interval before and after the therapist had attended a supervision session. The therapeutic bond was rated more favorably both by patients and therapists in the therapy interval following the supervision session. Such improvements occurred not only after specifically requested supervision but also after regular supervision, which indicates how supervision can positively affect the therapeutic process. M. J. Patton and Kivlighan (1997), in their study involving 75 counselors and 75 clients, found that the supervisee’s perception of the supervisory alliance was significantly related to the client’s perception of the counseling alliance. Lehrman-Waterman and Ladany (2001) focused on the evaluation process within supervision and could empirically demonstrate that effective goal-setting and feedback practices led to a stronger supervisory working alliance, enhanced trainee self-efficacy, and increased trainee satisfaction with supervision. A good working alliance defined by mutual agreement on the goals and tasks of supervision and a positive emotional bond has been shown to be positively related to trainee self-efficacy and a positive therapeutic working alliance (Efstation, Patton, & Kardash, 1990; Patton & Kivlighan, 1997).

In addition to the overall lack of research, no generally accepted theory of supervision exists in the field of psychotherapy. Each school of psychotherapy has its own idiosyncratic procedures and concepts (Schmelzer, 1997), and only recently have empirically grounded integrative concepts of supervision been elaborated (Falender & Shafranske, 2006). Bennett-Levy (2006) presented a cognitive model of skill acquisition differentiating declarative, reflective, and procedural elements of therapeutic skills. Newman (1998) and Milne (2006) discussed similarities and differences of therapy and supervision and stressed the need for further research in this area.

There are multiple reasons for this lack of empirical research: The necessity of evaluation might have been overlooked because the effectiveness of supervision was considered to be self-evident; the additional and nonreimbursable expense of participating in evaluation research might have been shunned; or participants might have hesitated to be evaluated, fearing negative results. Nevertheless, supervision of therapists in training is legally and ethically essential to ensure treatment integrity and patient welfare.

One way to enhance treatment quality is the monitoring of therapies with the help of adherence and competence scales like the Revised Cognitive Therapy Scale (Blackburn et al., 2001). This scale measures the competence of the therapists providing cognitive–behavioural therapy (CBT) and can be used as a tool in supervision, but it does not evaluate the quality of the supervision session itself. However, to ensure this, supervision itself needs to be the subject of evaluation and quality assessment. Moreover, this behavior therapy practiced in Germany cannot be fully equated with standard CBT, because behavior therapy in Germany, in addition to focusing on Axis I (of Diagnostic and Statistical Manual of Mental Disorders [fourth edition]; American Psychiatric Association, 1994) symptoms of psychopathology, quite regularly also works with interpersonal, cognitive, and emotional maladaptive patterns that maintain or moderate the patient’s symptoms or psychosocial problems. These comparatively broad goals are made possible through the German legal health insurance system, which finances even non-chronic depressive or adjustment disorders for up to 45 to 60 sessions of behavior therapy. In Germany, the term “verhaltenstherapie” (behavior therapy) is used broadly and covers multimodal behavior therapy (MBT), which allows the application of different intervention techniques (e.g., exposure, emotional restructuring, Socratic dialogue, interpersonal approaches; Zarbock, 2008). Considering the number of therapy sessions available and the broad range of goals and intervention techniques, it seemed quite plausible that the supervisors would expand the classic behavioral perspective of training and problem solving to more interpersonal and emotion-focused issues in the supervision session. Therefore, we expected that supervision conducted in Germany would place equal weight on clarifying and relationship issues. For this reason, an integrative model that broadens the perspective of CBT seemed to be most suitable as a basis for our evaluation research.

To account for this theoretical basis of our evaluation instrument, we, therefore, adopted Grawe’s model of general psychotherapy (Grawe, 1999), founded on empirical research, which entails a large groundbreaking meta-analysis of psychotherapy efficacy and effectiveness. Grawe (1999) defined therapeutic relationship, problem-solving, and clarifying motives (besides resource activation) as the main aspects of successful psychotherapy. Drawing on the idea of similarity between the therapeutic and the supervisory process (Milne, 2006), we hypothesized that Grawe’s three aspects of psychotherapy are essential for the supervisory process as well.
Successful supervision provides the supervisee with a supporting relationship to help solve treatment problems (e.g., choosing which technique to apply) and to clarify treatment-related issues (e.g., discover functionalities of symptoms or interpersonal behavior). Because Grawe’s model does not prescribe specific contents but highlights processes of supervision, it seemed usable to us as a metamodel. It allows the description of the supervisory processes conducted by different supervisors, who were estimated to be from different backgrounds and, therefore, to use a variety of different approaches encompassed within the large concept of German behavior therapy.

Previous research on supervision in the field of social work (Baer, 1998) also used Grawe’s model to operationalize the performance of supervisors. This was done through the application of an 18-item questionnaire incorporating the scales Relationship Perspective, Problem-Solving Perspective, and Clarifying Perspective. The impact of a good supervisory relationship on successful supervision is self-evident. Empirical research conducted with experienced supervisors showed a high percentage of statements “in which information, recommendations and proposals for solutions” were given. Even prominent therapists doing supervision “in general seemed to be strikingly schoolmasterly” (Auckenthaler, 1999). This approach to supervision could be regarded as predominantly emphasizing the problem-solving perspective.

A clarifying perspective can be considered as helping the supervisee focus on the case conceptualization or treatment and relationship problems by offering a new perspective (e.g., from the perspective of cognitive plans or rules or from hitherto neglected emotions). Furthermore, the training of the therapists in functional analytical reasoning (i.e., analyzing behavior as a function of antecedents and consequences) via guided discovery or facilitation of cognitive emotional self-exploration might be considered part of such a perspective.

The previously mentioned factor of analytically derived needs of the supervisee according to Frank and Vaitl (1986, 1987) can also be classified as belonging to one of the aforementioned three perspectives of Grawe: (1) relationship perspective (emotional support); (2) problem-solving perspective (enhancement of competence, case- and method-related guidance), and (3) clarifying perspective (open emotional space, activation, and intellectual challenge).

Following these thoughts, we wanted to develop an instrument to measure the process quality of supervision. The design of the instrument should also meet the following criteria: collecting data on the perspectives of supervisor and supervisee, respectively; time efficiency and user acceptance; and ability to include the patient’s perspective as well.

The already existing Stundenbogen für die Einzelpsychotherapie (STEP [Questionnaire for the General and Differential Single Therapy Session]; Krampen, 2002) provided us with an ideal model for the design of our own questionnaire. The STEP is a 12-item questionnaire designed to evaluate a single therapeutic session. There are two versions, each one meant to be completed by either the patient or the therapist. The STEP is based on Grawe’s three-component model of psychotherapy, and all three components were empirically replicated by means of scale and factor analysis (Krampen, 2002).

Matching the items of the STEP as much as possible, we constructed our version to evaluate supervision sessions (Stundenbogen zur Supervisionsbeurteilung; SSB [Questionnaire to Evaluate Supervision]), questioning supervisors and supervisees separately. More details are given in the Measures section.

Aim of the Study

The aim of this study was to show the descriptive power of an instrument based on the three dimensions of Grawe’s model for the supervisory process. In a second step, we wanted to test the hypotheses that the perspectives of the supervisee and the supervisor would correlate in a similar pattern and order as found by Krampen (2002) concerning the perspectives of patients and therapists. Krampen (2001) found that the Clarifying, Problem Solving, and Relationship subscales correlated highest with their respective counterparts between the perspectives. This means, for example, that in the intercorrelation matrix the correlation between both Relationship subscales (i.e., the therapist’s and the patient’s) were higher than all other observed correlations (e.g., between the Relationship subscale of the therapist and the Problem Solving subscale of the patient). If our hypotheses were confirmed, this could be seen as a validation of our thesis of structural similarity between therapeutic and supervisory process.

Furthermore, we wanted to explore the extent to which the perspectives of the supervisor and the supervisee share common variance and how a general sense of satisfaction with the supervision session can be predicted by our more detailed and theoretically based questionnaire (SSB) in order to better understand the supervisory process.
This study was also conducted to facilitate further empirically based steps in quality assurance of supervision in an MBT training setting. By proving the utilizability of our instrument, we hoped to advocate an implementation of the instrument in the evaluation of therapy training on a regular basis. We hoped that a detailed and probably also controversial discussion with our peers about the underlying models and assumptions of their specific approaches to behavior therapy could be encouraged through this measurement.

**Method**

**Participants**

In our study, 37 supervisors (psychologists and psychiatrists) and 90 supervisees treating 90 patients participated. The supervisors worked in postgraduate training institutions in Germany, which offer training for psychologists to become psychotherapists with behavioral therapeutic orientation. This training can take up to 5 years and includes more than 4,200 hr of different training components (e.g., psychiatric training in a hospital, self-experience, therapeutic skill training, case work, supervision).

Of the 37 supervisors, 18 were women and 19 were men. The mean age was 44.5 years ($SD = 7.8$). Supervising experience ranged from 1 to 27 years ($M = 11.6$ years, $SD = 5.7$). Of the 90 supervisees, 72 were women and 18 were men. They had an average of 38.9 months of training in psychotherapy (range = 2–78 months). Of the 90 patients, 62 were women and 28 were men. Patient diagnoses were classified as follows: affective disorder, 24%; phobic disorder, 40%; eating disorder, 6%; personality disorder, 14%; “other,” 16%. Fifteen supervisors supervised one case, four supervisors supervised two cases, 10 supervised three cases, three supervised four cases, and five supervised five cases. A mean of 18.1 ($SD = 13.9$) therapy sessions were supervised ($mdn = 15$, range = 1–58).

**Measures**

Because the SSB resembled the STEP, we obtained permission from the author. Krampen himself also checked this adaptation for accuracy. Finally, the layout of the SSB was adjusted to the STEP. The SSB consists of 12 items, of which five refer to the clarifying perspective, four to the problem-coping perspective, and three to the relationship perspective (Table I).

Table II and III show, besides the loadings of the factor analysis, the items of the Relationship, Problem Coping, and Clarifying subscales for the supervisee and the supervisor.

In addition to the SSB, the following three items about supervision satisfaction were included to test concurrent validation: “I am satisfied with the process of today’s supervision session”; “I am satisfied with my own contributions to today’s supervision session”; “I am satisfied with the contributions my supervisor/the supervisee made to today’s supervision session.” All items (i.e., the SSB and the satisfaction items) were coded on a 7-point Likert scale, with a theoretical range from 1 (not true at all) to 7 (totally true).

**Procedure**

Between October 2002 and April 2003, 150 SSB questionnaires were sent to behavior therapy-oriented supervisors throughout Germany. These questionnaire included three additional items that reflect the degree of satisfaction with the supervision session and further items about sociodemographic and profession-related data of the supervisor and supervisee as well as data of the supervised case. The questionnaires were distributed by the directors of training institutions, who had signaled their willingness to cooperate and who were contacted by Gerhard Zarbock during organizational meetings of the conference of training institutes from the German Association of Behavior Therapy. The dissemination of the questionnaires for the supervisors and supervisees was done before the...
start of a supervision session. The supervisors asked for supervisee consent to evaluate the ongoing session. At the end of the session, the supervisor and supervisee independently completed their respective versions of the questionnaire. The supervisee enclosed the completed questionnaire and the additional sociodemographic data in an envelope and sealed it. The supervisor then put this envelope together with his or her own questionnaire and sociodemographic data in a prestamped envelope and immediately mailed it to Gerhard Zarbock. This procedure was designed to guarantee the anonymity of both supervisor and supervisee.

Each supervisor was allowed to evaluate up to 10 supervision sessions, but each evaluated session had to be done with a different supervisee. Ninety

Table II. Questionnaire Items for Supervisees with Factor Loadings

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Item</th>
<th>Factor</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship</td>
<td>3: I have understood very well what my supervisor meant.</td>
<td></td>
<td>.84</td>
<td>−.07</td>
</tr>
<tr>
<td></td>
<td>6: Today the therapist got help and support from me.</td>
<td></td>
<td>.78</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>10: Today I felt that my supervisor understood me.</td>
<td></td>
<td>.64</td>
<td>.35</td>
</tr>
<tr>
<td>Problem Coping</td>
<td>2: Today I got practical advice to cope with my problems with the patient in therapy</td>
<td></td>
<td>.54</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>5: Today I learned about alternative treatment strategies.</td>
<td></td>
<td>.47</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>8: What I have experienced today will help me in the future to cope better with my difficulties in therapy.</td>
<td></td>
<td>.47</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>12: What I have learned today will help me in forthcoming sessions to act more productively.</td>
<td></td>
<td>.70</td>
<td>.40</td>
</tr>
<tr>
<td>Clarifying</td>
<td>1: During today’s supervision session some of my difficulties with the patient in therapy appeared in a new light.</td>
<td></td>
<td>.02</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>4: In today’s session I got more insight into my behaviour and my reactions as therapist.</td>
<td></td>
<td>.27</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>7: Today it became clearer to me what exactly I want to do in therapy.</td>
<td></td>
<td>.61</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>9: Today I succeeded in seeing the patient and the therapeutic process, respectively, more sophisticated.</td>
<td></td>
<td>.43</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>11: I could look at the patient and the therapeutic process, respectively, from a new perspective.</td>
<td></td>
<td>.12</td>
<td>.81</td>
</tr>
</tbody>
</table>

Note. N = 90. Factor loadings are based on principal component analysis and varimax rotation with Kaiser normalization. Higher loadings are shown in boldface.

Table III. Questionnaire Items for Supervisors with Factor Loadings

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Item</th>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship</td>
<td>3: The therapist understood very well what I meant.</td>
<td></td>
<td>.22</td>
<td>.71</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>6: Today I got help and support from the supervisor.</td>
<td></td>
<td>.06</td>
<td>.72</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>10: Today the therapist felt that I understood her/him.</td>
<td></td>
<td>.23</td>
<td>.83</td>
<td>−.01</td>
</tr>
<tr>
<td>Problem Coping</td>
<td>2: Today the therapist got practical advice to cope with her/his problems with the patient in therapy.</td>
<td></td>
<td>.17</td>
<td>.61</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>5: Today the therapist learned about alternative treatment strategies.</td>
<td></td>
<td>.01</td>
<td>.15</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>8: What the therapist has experienced today will help her/him in the future to cope better with her/his difficulties in therapy.</td>
<td></td>
<td>.36</td>
<td>.16</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>12: What the therapist has learned today will help her/him to act more productively in forthcoming sessions.</td>
<td></td>
<td>.65</td>
<td>.11</td>
<td>.39</td>
</tr>
<tr>
<td>Clarifying</td>
<td>1: During today’s supervision session some of the therapist’s difficulties with the patient in therapy appeared in a new light.</td>
<td></td>
<td>.70</td>
<td>.00</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>4: In today’s session the therapist got more insight into her/his behavior</td>
<td></td>
<td>.68</td>
<td>.17</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>7: Today it became clearer to the therapist what exactly she/he wants to do.</td>
<td></td>
<td>.65</td>
<td>.38</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>9: Today the therapist succeeded in seeing the patient and the therapeutic process, respectively, more sophisticated.</td>
<td></td>
<td>.80</td>
<td>−.02</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>11: The therapist could look at the patient and the therapeutic process, respectively, from a new perspective.</td>
<td></td>
<td>.74</td>
<td>.20</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note. N = 90. Factor loadings are based on principal component analysis and varimax rotation with Kaiser normalization. Highest loadings are shown in boldface.
questionnaires were returned, for a response rate of 60%, which can be regarded as satisfactory.

Data Analysis

The SSB questionnaires for the supervisee and the supervisor were analysed separately, focusing on their psychometric criteria (item difficulty, item discrimination index, Cronbach's alpha reliability). To explore the factorial structure of the data set, a factor analysis was used.

The SSB self-reports (of the supervisee) can be seen as expert ratings (of the supervisor) concerning the supervision session. The perspectives of both can be compared and related to each other. This is, according to Catell (1965), a combination of life record and questionnaire data. This combination fulfils the criterion of a bimodal clinical psychological process diagnostic (G. Krampen, 2002, p. 25). To investigate the interrelation of both perspectives as well as the interrelations between the different dimensions (i.e., clarifying, problem coping, relationship) within each perspective, correlational analyses were conducted.

To determine whether overall satisfaction with supervision can be predicted by questionnaire data, a regression analysis was conducted, using a stepwise procedure with the overall satisfaction as a criterion and the three questionnaire scales or single items from the three scales, respectively, as predictors.

Results

Psychometric criteria for both versions of the questionnaire are reported separately. Furthermore, there are reports on the three subscales and on summarized scales adding up all 12 items, reflecting an overall evaluation of the session (combined scales).

Reliability

The reliability of our data was assessed using internal consistency (Cronbach's alpha).

Perspective of the supervisee. The means of the five items of the Clarifying scale were all between 5 and 6, given a 7-point Likert scale with a theoretical range from 1 (not true at all) to 7 (totally true). The mean of 5.3 indicated an average slight to moderate agreement with the items, which were scored in a positive direction; higher scores indicated more reported experience of clarifying insights during the session.

The corrected item-scale correlations varied between .41 and .72, which was satisfactory. Cronbach's alpha of this scale was .77, which seems to be acceptable regarding the relatively small number of items.

The means of the four items of the Problem Coping scale varied between 4.5 and 6.1. Item 4 was the lowest scoring item referring to the acquaintance with new treatment alternatives. The corrected item-scale correlations ranged from .38 to .62. Cronbach's alpha reached .73. The means of the three items of the Relationship scale scored between 6.1 and 6.3. They had in comparison the highest means across all scales. The corrected item-scale correlations varied from .53 to .71. Cronbach’s alpha, despite the low number of items, was high in comparison: .78. The corrected item-scale intercorrelations of the combined scale ranged from .66 to .71, and because of the larger number of items, Cronbach’s alpha reached .86.

Perspective of the supervisor. With the Clarifying scale, means varied between 4.36 and 5.24. The corrected item-scale intercorrelations ranged from .53 to .62, and Cronbach’s alpha was .80. The Problem Coping scale means varied between 4.49 and 5.62. The corrected item-scale intercorrelations ranged from .36 to .50, and Cronbach’s alpha was .66. On the Relationship scale, means varied between 5.5 and 5.8. The corrected item-scale intercorrelations range from .42 to .60, and Cronbach’s alpha was .68. The combined scale had corrected item-scale intercorrelations ranging from .38 to .51, and Cronbach’s alpha was .83.

Concurrent Validity

To assess the concurrent validity of the SSB scales, three content-valid items of different aspects of satisfaction with the supervision session (satisfaction with course of supervision session, satisfaction with own contributions, satisfaction with the contributions of the supervisor or supervisee respectively) had been included. Table IV shows their intercorrelations.

Perspective of the supervisee. The highest correlation of the Relationship scale was found with satisfaction with course of supervision session and the lowest with satisfaction with own contributions. Satisfaction with the course of the supervision and satisfaction with the supervisor’s contributions correlated higher with the subscales of the SSB than satisfaction with own contributions. The correlational patterns can be interpreted in the direction of concurrent validity, because the SSB assesses mainly what the supervisee “got” from the supervisor.

Perspective of the supervisor. Within the perspective of the supervisor, the Relationship scale correlated highest with satisfaction with course of supervision, as it was for the supervisee. In addition,
as was expected, satisfaction with own contributions of the supervisor correlated higher with the SSB scales than satisfaction with the supervisee’s contributions, with the exception of the Relationship scale. Again, this correlational pattern can be interpreted in the direction of concurrent validity, because the supervisor’s own work and effort should structure the session.

Intercorrelations of the Scales and Exploratory Factor Analysis

Perspective of the supervisee. The intercorrelations for this scale were moderate to high. The exploratory factor analysis (principal components, varimax rotation) showed a factor solution, with one strong first factor explaining 42.74% of the variance. The first factor could be labeled as a Relationship scale factor with loadings >.60 for all items of this scale. Some of the Problem Coping scale items and one item of the Clarifying scale also loaded on this factor, indicating that, from the supervisee’s perspective, relationship enhancement and problem coping are not separated. A second factor, accounting for 11.64% of the variance, was marked by three items of the Clarifying scale, each loading >.60. The factor was labeled accordingly as clarifying. The relatively low loading of Item 7 on the Clarifying scale could be due to its formulation, “Became clearer to me what exactly I want to do in therapy,” which seems to have been understood as stressing the problem coping (i.e., “Want to do”) and not so much the “Became clearer to me” part, which was intended by us for the clarifying perspective. More details can be found in Table II.

Perspective of the supervisor. The intercorrelations of the items on this scale were moderate to high. The exploratory factor analysis resulted in a three-factor solution. A first factor was marked by all clarifying items loading >.60 and explained 35.24% of the variance. The second factor was marked by all relationship items, which loaded >.70 and explained 14.65% of the variance. The third factor explained 8.93% of the variance and was marked by two Problem Coping items, which loaded >.60.

Despite some items that were not loading on the scale as they were theoretically intended to do, the results of the explanatory factor analysis give some empirical support for the assumed three-dimensionality (i.e., clarifying, problem coping, and relationship) in the supervisory process from the supervisor’s perspective.

Correlations Between the Perspectives of the Supervisor and the Supervisee

The results of our psychometric analysis allowed us, at least to some extent, to correlate the perspectives of supervisor and supervisee. Because of the moderate to high intercorrelations of the subscales and the satisfactory magnitude of the Cronbach’s alpha of the combined scale, we started our explorative analysis by calculating the correlation between the supervisor’s and the supervisee’s combined scales.

The Pearson correlation was very low (.15) and was not significantly different from zero. This was a striking result. To further explore this matter, we computed subscale intercorrelations. However, the only significant correlation of the supervisor’s with the supervisee’s perspectives was in reference to the clarifying perspectives of both and showed only a very small $r = .24$. In a further step, we correlated the three additional satisfaction items of supervisor and supervisee. Here we found significant and small to moderate correlations (Table V and VI). Thus, the overall satisfaction of supervisor and supervisee with the supervision session correlated at .42, indicating a moderately satisfactory agreement between supervisor and supervisee.

Regression Analytic Exploration of General Supervision Satisfaction

To investigate the determinants of general supervision satisfaction, we analyzed which scales and which variables of our questionnaire could best predict overall satisfaction with the supervision
We used a stepwise procedure, which selects the first predictor with the largest criterion correlation and excludes selected predictors post hoc if the F value of a partial correlation coefficient drops below $p = .10$.

We found that the general supervision satisfaction of the supervisee was best predicted by the Relationship scale ($\beta = .59$) and to a smaller degree by the Clarifying scale ($\beta = .27$), with an adjusted $R^2$ of .57. When, in a second regression analysis, the items were entered as predictors, we found that Item 10 ("I felt that my supervisor understood me"), which was part of the Relationship scale, was the best predictor. Even with just this item as the only predictor, the $R^2 = .53$ was quite high, and the inclusion of three more items (Items 7, 6, and 1) only raised the $R^2$ to .60.

The results are similar regarding the perspective of the supervisor. The general supervision satisfaction of the supervisor was best predicted by the Relationship scale ($\beta = .52$) and to a smaller degree by the Clarifying scale ($\beta = .26$), with an adjusted $R^2$ of .41. As for the supervisee, in a further regression analysis, entering the items as predictors, we found that Item 10 ("Today the therapist felt that I understood him/her") was entered first in the regression model ($R^2 = .32$) and also held the highest beta weights in the successive steps, including further predictors.

The final model explained an $R^2$ of .47 and contained Items 10, 1 ("Today the therapist could see some of his/her difficulties with the patient in therapy in a new light"), and 3 ("The therapist understood what I meant very well").

### Discussion

The purpose of this study was to investigate whether the three perspectives of general psychotherapy (i.e., clarifying, problem coping, relationship) are useful in the evaluation of supervision and whether a questionnaire constructed relating to these dimensions could withstand firm psychometric inspection.

The psychometric analysis shows satisfactory results, given the small number of items for each of the subscales and the combined scale. The combined scale had a good Cronbach’s alpha of .87 for the supervisee and .83 for the supervisor. The concurrent validity was established by correlational analysis of the scales under investigation with three items of general satisfaction with the session (i.e., overall satisfaction, satisfaction with own contributions, satisfaction with other person’s contributions). From a psychometric point of view, the 12-item scales can thus be regarded as suitable instruments to assess overall process quality of the supervision session.

The explorative factor analysis and the psychometric properties of the subscales (i.e., Clarifying, Problem Coping, Relationship) give further empirical support for the distinctiveness of the subscales, but the moderate to medium intercorrelation between the scales also justifies a combined index of all items to describe the quality of the session. The few items that did not load as expected on the three scales seem to be tolerable because of the otherwise satisfactory scale characteristics. Furthermore, would a change of items violate the idea of having a nearly identical instrument for the patient and supervisor.

### Table V. Correlation Between Supervisors and Supervisees on SSB Subscales

<table>
<thead>
<tr>
<th>Supervisor</th>
<th>Clarifying</th>
<th>Problem Coping</th>
<th>Relationship</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarifying</td>
<td>.24*</td>
<td>-.00</td>
<td>-.09</td>
<td>.11</td>
</tr>
<tr>
<td>Problem Coping</td>
<td>.15</td>
<td>.16</td>
<td>-.02</td>
<td>.10</td>
</tr>
<tr>
<td>Relationship</td>
<td>.17</td>
<td>.14</td>
<td>.13</td>
<td>.16</td>
</tr>
<tr>
<td>Combined</td>
<td>.21</td>
<td>.11</td>
<td>-.04</td>
<td>.15</td>
</tr>
</tbody>
</table>

*Note. SSB = Stundenbogen zur Supervisionsbeurteilung (Questionnaire to Evaluate Supervision).

*p < .05. **p < .01.

### Table VI. Correlation Between Supervisors’ and Supervisees’ Satisfaction

<table>
<thead>
<tr>
<th>Supervisee satisfaction with:</th>
<th>Course of supervision session</th>
<th>Own contributions</th>
<th>Supervisor contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course of supervision session</td>
<td>.42**</td>
<td>.33**</td>
<td>.38**</td>
</tr>
<tr>
<td>Own contributions</td>
<td>.24*</td>
<td>.17</td>
<td>.29**</td>
</tr>
<tr>
<td>Supervisee contributions</td>
<td>.43**</td>
<td>.38**</td>
<td>.34**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
therapist dyad as well as the supervisee and supervisor dyad?

From a CBT perspective, it is possible to critically comment that the items of the SSB questionnaire are too “generic” to be really useful in CBT and even in MBT supervision. Because CBT and MBT are guided by specific techniques and concepts (adherence), which can be more or less adequately administered (competence), it is necessary to evaluate supervision according to the degree to which it fosters the supervisee’s competence in CBT or MBT. After viewing our results, especially the nonsignificant correlation between the perspectives of the supervisor and the supervisee (see later discussion), we have to admit that this kind of criticism may be justifiable. The quality of supervision could be defined as the degree to which the supervisory process can enhance the adherence and competence of the supervisee in applying specific strategies or techniques of diagnosing or therapy. It is possible that our questionnaire is confounded with consumer and provider satisfaction and does not only assess the quality of supervision.

Previously conducted research using our master questionnaire, the STEP, and evaluating the therapy session from the perspective of the patient and the therapist found correlations between the therapist’s and the patient’s evaluation of the same topic (e.g., assessing the relationship or the clarifying aspects of the session), which ranged from .44 (Clarifying) to .47 (Relationship) to .50 (Problem Coping) in various samples (Krampen & Wald, 2001). In our study, a correlation of this size was only visible between the global session satisfaction ratings of supervisor and supervisee \( r = .42 \). However, the more differentiated and theoretically based items of our 12-item questionnaire showed a nearly zero correlation between the assessments of supervisor and supervisee.

The means of the scales from supervisees and supervisors seem to be similar, even with a trend toward the supervisees judging the quality of the session slightly better. The puzzling result is that, even though with a similar questionnaire a correspondence between the perspectives of patient and therapist has been found (K. Krampen & Wald, 2001), our study failed to establish a similar correspondence between the perspectives of the supervisor and supervisee. This does not appear to be due to a sampling error reflecting a specific supervision style or training atmosphere, because we collected our sample from many different settings and not according to non-normal distributed scores. It seems, however, that despite being present at the same session and sharing similar components of the subjective evaluation process, supervisor and supervisee do not agree on what exactly happened in the supervision session. This gives rise to some doubt about the supervision processes in the training settings under investigation. Why are the supervisors, mostly experienced therapists, not able to detect what the supervisees are experiencing during the supervision sessions?

Our results replicated those of Efstation et al. (1990), who constructed parallel forms of the Supervisory Working Alliance Inventory (SWAI). The correlations between supervisee’s and supervisor’s perspectives on the different aspects of the supervisory process were amazingly low. The Client Focus subscale of the SWAI reflects the emphasis the supervisors placed on promoting the supervisee’s understanding of the client and thus resembles our Clarifying scale. The correlation between supervisor and supervisee was .21. The Rapport subscales, reflecting the supervisors’ effort to build rapport with the supervisees by supporting and encouraging them, correlated between the supervisors and the supervisees to .23.

Efstation et al.’s (1990) and our results resemble those of many of the studies that focus on the therapist–patient relationship. As Hill and Lambert (2004), p. 87) summarized, “Empirical research has found low correlations among the perspectives on various measures, supporting the view that the therapy experience varies for people in different roles in the therapy process.” Furthermore, it appears that therapists in training do not always tell the “truth of the case” in the supervision session but rather play a role to please the supervisor and to meet the expected demands placed on them by the training curriculum and the licensing procedure. Yourman (2003) labeled this effect as trainee non-disclosure induced by shame. In a qualitative study, Hess at al. (2008) found that nondisclosure in supervision of predoctoral interns was quite common and reported that even those interns with satisfying supervisory relationships withheld information from their supervisors out of fear of potentially negative evaluation.

Another possible explanation lies in the fact that supervision often relies on verbal reports. As far as we know, supervision by video or audiotapes is not done regularly and live supervision rarely happens at all. The process of case supervision by verbal reports is quite ambiguous and blurred and probably prone to strategies of social desirability and impression management by the supervisee. The fact that, as found in our study, the relationship dimension contributes most to overall session satisfaction of the supervisee and supervisor also supports this notion. As in psychotherapy treatments, a good relationship should be the basis of the supervisory process, and the importance of the supervisory bond is underscored...
by empirical findings whereby a strong need for emotional support is articulated by supervisees. However, for the acquisition of new skills, there also has to be, in a Piagetian sense, a challenge and subsequent accommodation of cognitive and emotional schemas of the supervisee, which in itself may cause some distress. If the supervisor is driven too much by his or her conscious or unconscious plan to be liked by the supervisee, the effort of challenging dysfunctional therapeutic behavior and stimulating the acquisition of new therapeutic and interactional skills may fall short. A study by Foster et al. (2007), looking at the supervisor–supervisee relationship from the perspective of attachment theory, also found only an insignificant relationship between the competence ratings of the supervisee (self-rating) and the supervisor (rating supervisee’s competence). The authors interpreted their findings in a similar direction as we are doing with ours, by acknowledging that “although a supportive relationship may facilitate growth by providing the supervisee with motivation and an optimistic self-evaluation, it may also prevent the fledgling supervisee from full awareness of skill deficits and areas of weakness.” To “cure” this problem, the obligatory use of videotapes of the patient sessions in supervision could be discussed. If videotapes are a “must-have” in supervision, the supervisee will get used to it and the real story can be heard for the benefit of the treated patient.

Nevertheless, another point of view is possible by assuming that the experience of supervisor and supervisee is really different, as are the experiences of patients and therapists. This would require more research into understanding the internal processes, opinions, judgments, expectations, and strategies of the supervisee and the supervisor on a deeper level. Qualitative research could be a first step toward a broader understanding of these issues (M. Q. Patton, 1990). Another conclusion could be that more structured procedures in the supervision session are needed. This goal could be realized by using adherence and competence scales on a regular basis to discuss audio or videotapes of the supervisee in the supervision session.

The results of our study imply that further research is needed. However, they also encourage us to invest as much energy, vigor, and thought in supervision research as has been invested in psychotherapy research.

Limitations. There are some limitations to this study. First, there was not a one-to-one match of supervisees and supervisors (90 vs. 37, respectively), and the statistical independence of the supervisor sample may have been violated because some supervisors judged more than one supervisee. Second, there was no control over the informal talk about the study between the supervisor and supervisee before completing the questionnaire. However, such variables should have led to more similarity of judgment, not less as we found. Third, our study was conducted in training settings, whereby supervisees were involved in a curriculum to become licensed multimodal behavior therapists. Fourth, because the participants of our study were supervisees in training and not experienced therapists, the need for emotional support when starting to practice psychotherapy could have overruled their need for intellectual challenge. Fifth, the study was based on self-ratings by supervisors and supervisees, but no objective assessment of the actual quality of the supervision session was acquired. Sixth, the supervisors probably preferred different models and strategies of (behavior) therapy and the supervisees were applying various different techniques in the supervised sessions, so that this variation adds to the variance without being sufficiently controlled or explicated. Seventh, the lack of reversed-scored items could have facilitated a positive response bias. Eighth, in an effort to minimize the workload of the supervisors and supervisees, no additional instruments to assess the convergent validity (e.g., SWAI; Efstation et al., 1990) were included; therefore, potential overlaps between the SWAI and the SSB could not be investigated.

Recommendations. An important step for future research would be to distinguish among verbal report, audiotape and videotape supervision, with the hypothesis that, by using videotape supervision, the results might be more convergent. Furthermore, a potential future direction could be the development of a coding version of the SSB to include an independent third-person perspective of the supervisory process.

A longitudinal study design using hierarchical linear modeling (Kline, 1998) would offer the opportunity to examine the therapist as a nested variable and to address the predictive validity of the SSB. Such designs could investigate the question of how supervision processes predict subsequent satisfaction with supervision, treatment process, or treatment outcomes.

Acknowledgements

We are grateful for the cooperation and contributions of the supervisors and supervisees of the DVT (German Association of Behavior Therapy) who participated in our study. We thank the unknown reviewers for their helpful suggestions. We also thank Johanna Grey, MSc, and Nancy Novick for their linguistic help.
References


