Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: https://www.tandfonline.com/loi/tpsr20

Coder Agreement Using the German Edition of Luborsky's CCRT Method in Videotaped or Transcribed RAP Interviews

Brigitte Zander , Michaela Strack , Manfred Cierpka , Günther Reich & Herrmann Staats

To cite this article: Brigitte Zander , Michaela Strack , Manfred Cierpka , Günther Reich & Herrmann Staats (1995) Coder Agreement Using the German Edition of Luborsky's CCRT Method in Videotaped or Transcribed RAP Interviews, Psychotherapy Research, 5:3, 231-236, DOI: 10.1080/10503309512331331336

To link to this article: https://doi.org/10.1080/10503309512331331336

Psychotherapy Research 5(2) 231–236, 1995

CODER AGREEMENT USING THE GERMAN EDITION OF LUBORSKY'S CCRT METHOD IN VIDEOTAPED OR TRANSCRIBED RAP INTERVIEWS

Brigitte Zander
Michaela Strack
Manfred Cierpka
Günther Reich
Herrmann Staats
University of Göttingen

In a reliability test of the German edition of the CCRT method, the amount of agreement for the coded contents of transcripts is compared with that from video tapes. The results show only a moderate agreement in the various stages of evaluation. Transcribing video episodes demands a great deal of work and does not result in increased reliability. The results suggest that the categories of contents should be reformulated.

INTRODUCTION AND AIMS

The development of methods for the quantitative diagnosis of interpersonal relationship patterns should help improve the objectivity and transparency of clinical judgment and provide valid content categories for basic research. The Core Conflictual Relationship Theme (CCRT) method (Luborsky, 1977; Luborsky & Crits-Christoph, 1990; for the German edition called *Zentrales Beziehungskonflikt Thema* (ZBKT), see Luborsky & Kächele, 1988) is a standardized analysis of the contents of interpersonal narratives which works out repetitive patterns of wishes and responses of one person in interaction with relevant others, similar to a clinical assessment (cf. Soldz, 1993).

Reports of relationship episodes are examined for three components derived from psychodynamic and interactional theory: the subject's wish (need, intention), which produces the response of the other and then again the response of the self. The narrative material is collected either from records of therapy sessions or from so-called RAP interviews (Relationship Anecdotes Paradigm; Luborsky, 1990b; *Beziehungsepisoden* Interviews; Dahlbender, Torres, Reichert, Stübener, Frevert, & Kächele, 1992). In this specialized form of interview the client is asked to relate at least ten subjectively important and concrete events ("narratives") experienced with different people at any time of his or her life. These narratives provide the basis of our investigation.

The authors were assisted by J. Biskup, H. Homburg, S. Krannich, K. Ratzke, and L. Seide.

Correspondence regarding this article should be addressed to Brigitte Zander, Center for Family Therapy, Department of Psychosomatics and Psychotherapy University of Göttingen, Humboldtallee, 37073 Göttingen, Germany.

The following short contribution[1] reports on a reliability study carried out for the German edition of the CCRT method, which pursued four criteria:

Amount of agreement: First it was necessary to estimate the general degree of coding agreement for the German edition of this method. The amount of agreement is determined for the identification of the relationship episodes and for the grouping of the components (wish, response of other, response of self) in eight clusters each (see German edition of the manual; Luborsky, assisted by Albani & Eckert, 1991).

Type of presentation: The primary aim of the investigation was to compare the coding agreement of transcribed and videotaped interviews, which is entirely wanting. If transcription cannot be shown to be superior, then it will no longer be necessary to employ the time-consuming method of transcribing all narrative material.

Interdependence of narratives: Usually the method is used for therapy transcripts and interviews for relationship episodes: several episodes by the same narrator were coded consecutively, so that it can be assumed that the narratives are not independent of each other. If a similar agreement could be achieved for single episodes, each of which originate from a different subject, then this method could be tested for the description of relationships following shorter contacts, for example, after a case history. For this question we process two studies with different amounts of narrative dependence.

Data preselection: The fourth criterion was one of method. In the American reliability study of the CCRT method (Crits-Christoph, Luborsky, Dahl, Popp, Mellon, & Mark, 1988; Crits-Christoph, Luborsky, Popp, Mellon, & Mark, 1990) only those relationship episodes which had been unanimously identified by several coders were accepted for coding. The reason for this preselection was probably to obtain higher agreement values. Our study analyzes the complete material and its accordingly selected parts.

METHODS

Design: Two studies were carried out to investigate how far the coding agreement follows the interdependence of narratives. The same 6 coders took part in both studies. All had been trained in either systemic, analytic, or humanistic psychotherapy.Their professional experience varied from 1 to 17 years and they had between 3 and 6 months experience with the CCRT method.

The material for study I consisted of both video and transcript versions of 6 RAP interviews with nonclinical volunteers (female students). Each interview contained between 10 and 13 narratives. As material for study II, only the first narrative was taken from each of 65 RAP interviews. The subjects were female students, too.

To avoid confounding the quasi-experimental variable (video vs. transcript) with the groups of coders, a cross-validation design was employed: the coders were divided into two groups and each received half the material (so-called blocks) in the other type of presentation. Whether the type of presentation affected the agreement of the 3 coders can only be inferred from the replication of this effect in the other group with reversed blocks. All relationship episodes were coded on the basis of the German CCRT manual (Luborsky, assisted by Albani & Eckert, 1991).

Evaluation: The agreement in the identification of relationship episodes is

[1]A fuller account, in German, is available from the author.

expressed by the relative frequency (percent) with which an arbitrary second coder would make the same identification (confirmation probability; Po, given by Fleiss; see Bortz, Lienert, & Boehnke, 1990). The agreement in the grouping of the components *wish, response of other,* and *response of self* in the 8 content clusters is defined by Cohen's kappa, as modified by Fleiss. This was done for all data and also for only those relationship episodes that were identified by all 6 coders.

RESULTS

Amount of agreement in the identification of relationship episodes. The coders identified maximally 83 relationship episodes within the 65 interaction narratives of study I and maximally 92 episodes within the 65 interaction narratives of study II. However, only 71% in study I and 63% in study II were recognized by all 6 coders (study I, Po = 92%, study II, Po = 89%).

If a narrative contains only one episode (74% respect. 48% of the episodes) then there is a very high confirmation probability that it is identified (Po 97–98%). In narratives where 2 episodes are identified, Po fell from 89–94% for the first to 75–66% for the second episode of the same narrative (for the third episode within one narrative the confirmation probability sinks to a Po of 20%).

Type of presentation in the identification of relationship episodes: In study I the video presentation showed itself to be superior in both combinations of interviews and coder groups. In the first 3 interviews, group 1 achieved from video a Po=97% in comparison with group 2 from transcript a Po=93%. While in the second interview block, group 2 achieved with video a Po=93% better than group 1 with transcripts Po=84%. However the advantages of video could not be repeated in study II. Here considerable differences for both material blocks were found. There was more agreement in the episode identification in the first 32 narratives in both the video (group 1, Po=96%) and the transcript (group 2, Po=93%) than in the 33 narratives of the second block (video, group 2, Po=82%; transcript, group 1, Po=84%).

Amount of agreement in the cluster grouping: If only the episodes are examined which were unanimously recognized by all 6 coders the mean agreement over all 3 components and both presentation modes for the clusters choice is with <Po>=.553 and <kappa>=.439 in study I only insignificantly higher than on the basis of all relationship episodes (<Po>=.521, <kappa>=.405).

This difference is larger in study II (only episodes identified together <Po>=.534, <kappa>=.434, all episodes: <Po>=.473, <kappa>=.363). Overall, the results of both studies show relatively low agreement values.

Type of presentation in the cluster grouping: In study I, it seemed that the video presentation of the first 3 interviews led to a higher agreement (kappa=.44–.58, group 1) than the transcript presentation (kappa=.38–.47, group 2). This relationship was reversed for the second interview block: the coders of group 1—now in transcript—were more in agreement (kappa=.35–.48) in comparison with group 2 in video (kappa=.14– .52). Therefore, no presentation effects were found but coder effects. The cross-validation design showed its controlling effect. Only in the codings of reactions of subjects could a slight advantage of the video presentation be observed in both blocks (kappa=.45–.52 in contrast to kappa=.38–.48).

It is striking that the varying agreements in both interview blocks was caused

chiefly by the coding of the wishes (maximum range of kappa=.14 to .58; notice that even a kappa=.14 has only 1% chance level in our data).

In Study II, the video presentation shows itself to be superior in the cluster grouping of the wishes in both narrative and coder group combinations (kappa=.26–.52 vs. wish component in transcript kappa=.21–.31). No superiority can be observed, as in study I, in the subject reactions. Here also there are considerable differences between both narrative blocks.

DISCUSSION

The assumption found in the literature that coding from transcript results in a higher reliability could not be supported by our findings. Study I indicated that there was some advantage in using video presentation in the identification of relationship episodes and the cluster grouping of reactions of subjects. But this could not be repeated in study II, where the video performance led to a higher agreement in the coding of the wishes which are usually felt as the most difficult component to be coded. At least the transcription of relationship episode interviews with nonclinical subjects is no longer essential for the reliability of the coding process.

This finding should not conceal the general weakness of the CCRT method in producing reliable results. For example the confirmation probability of around 90% in identifying relationship episodes means that every tenth relationship episode was not recognized by a parallel coder and so could not be included into the assessment, which closely followed the Luborsky manual. The core conflictual theme gained from relationship episode interviews or therapy dialogues are therefore based on different underlying episodes.

Additionally, the reported confirmation probability should only be interpreted as the upper limit of agreement because of the special kind of narratives provided in the framework of RAP interviews in which the 'minimum identification' of one episode per narrative seems much easier compared with those in a continuous therapy dialogue. A decrease in the identification agreement was shown where there were several episodes for each narrative (see a similar observation from Crits-Christoph and colleagues, 1988), and especially in the 'first narratives' used in study II. A high proportion of these first narratives contained several episodes. In narratives which are less structured than those required in the RAP interview, an even lower identification agreement should be found.

The assignment of the components *wish, response of other,* and *response of self* to 8 content clusters had a moderate confirmation probability of just 50% and increased only insignificantly after selection for the consistently identified episodes. The kappa coefficients found in both studies just reach the grade moderate (see Landis & Koch, 1970); though there were single values below .20. Overall, this reliability estimate is consistent with the values gained from transcribed therapeutic sessions in a small German study by Guitar-Amsterdamer, Stähli, Schneider, & Berger (1988). In our experience, the chief problem of the method seems to be the coding of the relationship *wishes.* Our data did not continuously show the lowest agreement values, but their largest range.

However, the greatest variances resulted from the coder groups and the interviewed subjects. The agreement differences caused by the material itself (the block-effects) can partly be understood by the clarity of the narratives. However, we were more concerned with the differences between the coder groups. Attempts to

explain them by a greater variability of opinion between experienced clinicians could not be upheld after examining single-group values.

The CCRT method tries to render the core conflictual relationship pattern, during its final formulation stage, insensitive to disagreement in details by means of a cumulation of at least 10 relationship episodes. However, it remains open as to whether the agreement deficits found in the preceding evaluation steps are actually compensated or whether the diagnosticians move further from each other during an interview coding owing to the schema–self-perpetuating process (see, e.g., Lewicki, Hill, & Sasaki, 1989, for this theory). To gain empirical criteria for this decision, the core conflictual relationship theme was formulated for the 6 interviews from study I. In fact, the mean agreement for a CCRT cluster with $<Po>=70\%$ is higher than that of 50% on the episode level; although the kappa coefficient of .44 agrees with the former. The differences in the coder groups and the interview blocks can also be found again in the CCRT.

What consequences can be drawn from the only moderate reliability of the method? The ambitious claim that the CCRT method ". . . behaves much like experienced psychodynamic clinicians do in their usual inference processes in formulating transference patterns, although it formalizes their principles of inference processing" (Luborsky, 1990a, p. 2), is certainly limited if the described weakness of the method could only be counteracted by recommending an intensive common training for the coders. Instead of this we prefer to improve of the content categories. By this we do not only hope to increase the method's reliability, but also to attain a better correspondence to the diagnostician's mental processes.

REFERENCES

Bortz, J., Lienert, G. A., & Boehnke, K. (1990). Verteilungsfreie Methoden in der Biostatistik. Berlin: Springer.

Crits-Christoph, P., Luborsky, L., Dahl, L., Popp, C., Mellon, J., & Mark, D. (1988). Clinicians can agree in assessing relationship patterns in psychotherapy. The Core Conflictual Relationship Theme Method. *Archives of General Psychiatry, 45,* 1001–1004.

Crits-Christoph, P., Luborsky, L., Popp, C., Mellon, J., & Mark, D. (1990). The reliability of choice of narratives and of the CCRT measure. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference* (pp. 93–101). New York: Basic.

Dahlbender, R. W., Torres, L., Reichert, S., Stübener, S., Frevert, G., & Kächele, H. (1992). Die Praxis des Beziehungsepisoden—Interviews. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse 39,* 51–62.

Guitar-Amsterdamer, H., Stähli, R., Schneider, H., & Berger, E. (1988). Können Komponenten konflicktiver Beziehungsmuster in einem psychotherapeutischen Gespräch reliabel identifiziert werden? In L. Luborsky & H. Kächele (Eds.), *Der Zentrale Bezichungskonflikt* (pp. 60–78). Ulm: PSZ Verlag.

Landis, J. R., & Koch, G. G. (1970). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Lewicki, P., Hill, T., & Sasaki, I. (1989). Self-perpetuating development of encoding biases. *Journal of Experimental Psychology: General, 118,* 323–337.

Luborsky, L. (1977). Measuring a pervasive psychic structure in psychotherapy. The core conflictual relationship theme. In N. Freedman & S. Grand (Eds.), *Communicative structures and psychic structures* (pp. 367–395). New York: Plenum.

Luborsky, L. (1990a). The early development of the core conflictual relationship theme idea. In L. Luborsky & P. Crits- Christoph (Eds.), *Understanding transference* (pp. 1–11). New York: Basic.

Luborsky, L. (1990b). The relationship anecdotes paradigm (RAP) interview as a versatile source of narratives. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference* (pp. 102–113). New York: Basic.

Luborsky, L. assisted by Albani, C., & Eckert, R. (1991). Manual zur ZBKT-Methode (Übersetzung und deutsche Überarbeitung mit Ergänzungen der Ulmer ZBKT- Arbeitsgruppe). Universität Ulm, Abteilung Psychotherapie.

Luborsky, L., & Crits-Christoph, P. (Eds.) (1990).

Understanding transference. The CCRT method. New York: Basic.

Luborsky, L., & Kächele H. (Eds.) (1988). *Der zentrale Beziehungskonflikt. Ein Arbeitsbuch.* Ulm: PSZ Verlag.

Soldz, St. (1993). Book review: *Understanding transference: The CCRT method.* Luborsky, L. & Crits-Christoph, P. New York: Basic Books, 1990. *Psychotherapy Research, 3(1),* 69–75.

Received February 19, 1993
Revision Received September 26, 1993
Accepted November 5, 1993