The Algorithmically Structured Systematic Exploration of Subject’s State of Mind: II Reliability and Construct Validity

MARIE-NOËLLE LE MER AND SERGE STOLÉRU

ABSTRACT

Objective: To investigate the reliability and validity of a newly developed semi-structured interview, the Algorithmically Structured Systematic Exploration of Subject’s State of Mind (Assess_Mind). Method: Twelve scales have been developed and used to rate 86 Assess_Mind interviews conducted with patients undergoing in vitro fertilization treatment. For each rating scale, we assessed interrater reliability, interviewer effect, and construct validity. In addition, a factor analysis of scales was performed. To study the validity of scales and of the factors yielded by factor analysis, patients were assessed with a psychopathological battery including the Child Project Questionnaire, the State-Trait Anxiety Inventory, the Perceived Stress Scale, and the Ways of Coping Checklist. Results: For 11 of the 12 scales, interrater reliability was fair to excellent. There was no significant interviewer effect affecting any of the scales. In addition, the validity of nine of these 11 scales was demonstrated by their correlations with questionnaires measuring similar constructs. Conclusion: Nine of the Assess_Mind scales appear to have adequate psychometric properties. Copyright © 2007 John Wiley & Sons, Ltd.

Key words: interview, psychological, psychometrics; in vitro fertilization, emotions, motivation

INTRODUCTION

Although important advances have been made in the last two decades in improving the standardization and the reliability of interview schedules, in particular regarding diagnostic interviews, methodological problems remain to be solved (Regier et al., 1998). Thus, in a series of landmark articles (Cox et al.,
1981a; Cox et al., 1981b; Hopkinson et al., 1981; Rutter and Cox, 1981; Rutter et al., 1981) it has been shown that some interview techniques were optimal to obtain factual information, while others were better adapted to elicit the expression of feelings. Although active detailed questioning and probing were associated with the collection of better-quality factual information, sympathetic statements from the interviewer were one of the techniques most often followed by the expression of feelings. Not only the nature of information (e.g. factual versus emotional) is influenced by specific interview techniques, but also the range or scope of information that may be collected. Closed questions usually focus on specific items while open questions leave much more freedom to respondents, which may elicit the expression of unexpected, but important, information. Thus, depending on the specific objective of the clinician or the researcher, different interview styles will be optimal.

While it is essential to maintain high standards of reliability, we may also be concerned that strict constraints imposed by some kinds of interview schedules may induce some loss of information, in particular regarding the expression of feelings. On the other hand, if open questions, sympathetic statements, and/or interpretations are used (Cox et al., 1988), we may be concerned that reliability, in particular, inter-interviewer reliability, is likely to be compromised. The level of variance between interviewers has been shown to be higher for unstructured than for structured interviews (Saghir, 1971; Miller, 2001). As demonstrated by Kerlinger (1986), high reliability of a measurement instrument is a necessary but not a sufficient condition for the validity of data. Therefore, if a method is not reliable, the assessment of its validity is not meaningful. Thus, researchers seem to be confronted with a dilemma: either to use unstructured interviews adapted to collect data on emotions, but in a potentially unreliable manner, or to use structured interviews with potentially adequate reliability and validity, but less adapted to collect data on emotions. To contribute to the resolution of this dilemma, based on psychoanalytical concepts we have developed a new research interview, the Algorithmically Structured Systematic Exploration of Subject’s State of Mind (Assess_Mind). More specifically, our purpose has been to design a research interview that would meet the following criteria:

- in terms of what it investigates, to be as open as possible as to the nature and content of information it could collect
- to meet usual criteria of scientific instruments, that is, reliability and validity.

The purpose of the Assess_Mind is to provide clinical researchers with a detailed and extensive “map” of the phenomenology of a subject’s mental state (Chessick, 2002). By “phenomenology,” we refer to those aspects of mental state that are either consciously experienced or that are accessible to consciousness upon questioning within the interview situation. Examples of the latter aspects include episodic memories. The Assess_Mind comprises two modules: a specific interview format, functioning as a data collection instrument, and a set of rating
scales that are described hereunder. As regards the interview format, five areas, called “functional registers” are sequentially investigated: affects in general; anxieties; desires; memories; and associations of ideas. Then, various dimensions of this phenomenological experience are quantitatively assessed through the rating scale module.

As it does not vary with the topic of the study, the interview format module is fixed. By contrast, the rating scale module is variable: it is composed of sets of scales that have been specifically developed in relation with the various topics investigated with the Assess_Mind. (Stoléru and Le Mer, 2007) For instance, the Assess_Mind has been developed in the context of a study of the psychological dimensions of in vitro fertilization (IVF) treatment. In the present study, we have developed scales that were deemed relevant to this topic, for example a scale assessing the intensity of the “Wish to Have a Child.”

In a previous paper we have described the development and the algorithm to administer the Assess_Mind. The objective of the present paper is to report the psychometric properties of the Assess_Mind, as used in the context of a study of the psychological dimensions of IVF. This will be done by examining the interrater reliability and the construct validity of the 12 rating scales developed in the study. The predictive validity of the instrument, that is, the relation between ratings and IVF outcome, will be presented in a subsequent paper.

METHOD

Patients

The study was approved by the local ethics committee, and written, informed consent was obtained after complete description of the study to the subjects. Eligibility criteria were: age less than 38 years for women; fluency in French; belonging to a Western culture (because of wide cultural variations in the psychological context of infertility); and gametes available from both spouses (i.e. functional ovaries and no azoospermia) because oocyte or sperm donation create distinct psychological situations. Fifty-seven women and 29 men, including 21 couples, treated through IVF in a private clinic, participated in the study. Men declined to participate more often than women did. This difference may have been related to the fact that the women were hospitalized on the eve of the day of oocyte retrieval (OR) and thus were more available. For female patients, the means (SDs) for age, years of education, and years of infertility were 32.4 (3.1), 14.9 (3.2), and 5.7 (3.7) years, respectively. The corresponding figures for men were 33.8 (4.3), 14.7 (3.0), and 5.0 (3.0) years, respectively. Fourteen women (24.6%) and five men (17.2%) had already children.

Assess_Mind Interview

As mentioned above, five registers were sequentially explored. Twelve scales, described in detail hereunder, were developed:
wish to have a child
negative affects associated with the project to conceive a child
negative affects associated with the IVF procedure
negative affects associated with infertility
negative affects associated with problems not directly related to the current IVF trial
positive affects associated with the project to have a child
vulnerability to psychosomatic disorganization
expression of emotions
positive episodic memories associated with the project to conceive a child
negative episodic memories associated with the project to conceive a child
husband’s support and concern for wife
representational activity related to IVF.

For each scale, the final rating was based on the frequency and strength of statements relevant to the operational definition of the scale. Here, “frequency” refers to the number of registers where statements relevant to the scale being rated were noted. The operational definitions of the scales and the rating procedures were specified in a manual (Stoléru and Le Mer, unpublished).

For each scale, the procedure was the following.

- In each register, every relevant statement was identified.
- A strength or intensity – “low,” “medium,” or “high” – was assigned to each of these statements.
- The final rating, which ranged between 1 and 9, was assigned using an algorithm taking into account the intensity and frequency of relevant statements.

Raters were blind as to measures derived from the other instruments used in the study and presented below.

Other Psychological Measures

Subjects completed a sociodemographic questionnaire investigating items potentially related to the psychological aspects of IVF and/or to its outcome, for example age and number of children. A medical questionnaire focused on items potentially associated with the outcome of IVF, for example length of infertility. A series of psychological questionnaires was presented and measures from these instruments were correlated with ratings of the Assess_Mind to assess the construct validity of the scales. We presented the Child Project Questionnaire (CPQ), a self-report questionnaire previously developed and validated by our group, that is used to measure specific psychological factors that may contribute to the etiology of infertility (Stoléru et al., 1993; Stoléru et al., 1997). The CPQ covers three domains: (i) the wish to have a child, but also anxieties related to
Reliability and validity of assessing the mind

the prospect of having a child/becoming a parent; (ii) qualitative and quantitative aspects of sexual life; and (iii) family history related to procreation. Two types of score are derived from the CPQ: (i) a priori scores, derived on a conceptual basis by summing items related to the same dimensions; these are the scores of Anxieties Associated with Prospective Parenthood (AAPP), the score of Difficulties in Family History related to Procreation, and the score of Sexual Difficulties; (ii) factor scores, derived from factor analysis. Separate factor analyses have been performed for women and men undergoing IVF treatment and have yielded two interpretable factors for women and one for men (Stolér et al., 1997). In both women and men, Factor I has been interpreted as a measure of the Wish to Have a Child. Scores of women on Factor II, interpreted as Perceived Marital Harmony in the Project to Conceive, were predictive of success of the fertilization step of IVF (Stolér et al., 1997).

The State and Trait Anxiety Inventory (STAI) (Y form) (Spielberger et al., 1983) is a widely used self-report questionnaire comprising a form intended to measure anxiety as a personality trait, and another that focuses on the current level of anxiety, that is, state anxiety. As the IVF procedure is potentially stressful, the Perceived Stress Scale (PSS) (Cohen et al., 1983) was used to evaluate the intensity of stress perceived by the patients within the last month. The Ways of Coping Checklist (WCC) (Vitaliano et al., 1985) has been designed to measure to what extent subjects use various coping strategies when facing stressful life events. Five factors have been identified in a validation study conducted on a French sample (Paulhan et al., 1994):

- problem-focused coping
- wishful thinking and avoidance of problem
- seeks social support
- positive reappraisal of stressful situation
- blame directed against self.

The Neonatal Perception Inventory (NPI) (Broussard, 1974) is intended to measure parental perceptions of their newborn baby as compared to their concept of the average infant. The inventory consists of two forms, the “Average Baby” and the “Your Baby” forms. For the infertile patients of the present study, only the “Average Baby” form was presented before embryo transfer (ET) to test the hypothesis that anticipated problems in the newborn would be associated with increased anxiety, a factor that may contribute to IVF failure.

The participants’ experience of infertility, their counselling and support needs, their experience of the IVF procedure, and of medical service provision and quality were assessed using questionnaires developed by Edelmann and Connolly (1987) and by Mahlstedt et al. (1987). A series of scores were derived from these questionnaires by summing answers on items measuring specific dimensions of the experience of infertility, of medical examinations, and of IVF treatment. These scores were:
negative affects associated with infertility
negative affects associated with medical examinations performed before the current IVF trial
negative affects associated with treatment performed before the current trial
negative affects associated with initial steps of IVF (decision to engage in trial, phase of ovarian stimulation)
negative affects associated with examinations and treatments performed during current trial
level of criticism of current medical care and clinic
perceived lack of support from others
interference between trial and other activities
summary score of negative affect associated with current trial (infertility, IVF)
global summary score of negative affects experienced since beginning of current trial, whether or not these affects were related to the trial.

Finally, the Social Desirability Scale (Crowne and Marlowe, 1960) was used to test the hypothesis that some of the above measures might be influenced by social desirability.

Procedure

The gynecologist invited all consecutive eligible couples to participate in the study. To encourage patients to express themselves freely, subjects were assured that the data would be kept anonymous and that questionnaire and interview information would be kept separate from clinical management. Participants were then provided by the gynecologist with a series of STAI and PSS forms, to be returned in a stamped pre-addressed envelope to the research team. Patients were requested to complete the STAI Trait form two days before OR, and to fill in the STAI State form on seven consecutive days, beginning two days before OR. The PSS was to be completed on the day before OR. In addition, on the day before OR, patients were interviewed with the Assess_Mind and the questionnaires mentioned above (see “Other psychological measures”) were administered. If embryos were subsequently obtained in vitro, patients were interviewed again and filled in a second CPQ four or five days later, a few hours before ET. Assess_Mind interviews were transcribed for subsequent rating.

Statistical Analysis

Assessment of interrater reliability
For each scale, interrater reliability was assessed by computing the intraclass coefficient of correlation (ICC) (Haggard, 1958) between ratings (ranging from
1 to 9) assigned independently to transcripts by two raters. Eighteen interviewers participated in data collection, and two of them (MNLM and SS) rated the transcripts. To eliminate a possible bias affecting the rating of interviews that had been conducted by one of the raters, we first assessed the interrater reliability of the ratings of interviews that had not been conducted by a rater. Then we performed a second analysis on the ratings of interviews conducted by one of the raters.

Assessment of interviewer effect
In addition to between-subjects variance, two sources of variance may influence ratings: the first is due to the interviewer; the second to the rater. By “interviewer effect,” we refer to the possibility that the same subject might give different answers to two different interviewers. To identify a potential interviewer effect, a useful strategy would be that the same subjects be interviewed by two different interviewers within a short interval. This was not possible because of technical and ethical problems: on the one hand, the period of IVF treatment is characterized by very rapid psychological changes and the Assess_Mind has been constructed to be sensitive to change; on the other, it seemed to us inappropriate to invite these couples – already undergoing several time-consuming and often stressful medical investigations – to participate in two relatively long (75 min) interviews separated by a short delay.

Nevertheless, we attempted to analyze the potential influence of interviewers by studying two sets of interviews conducted by two different interviewers (set A and set B) who had conducted enough interviews for statistical analysis to be performed. For each scale, we used the Mann–Whitney test to compare the ratings of these two sets. No systematic factor had influenced the assignment of a particular interviewer to a particular patient, so that this analysis was meaningful.

Assessment of construct validity
To assess the construct validity of the scales, we performed correlational analyses between the ratings and a series of other psychological scores measuring variables similar to the scales. First, a priori hypotheses were formulated about the sign and magnitude of all coefficients of correlations between the various Assess_Mind scales and the scores on the other psychological instruments. For some of these other measures, we hypothesized significant correlations with one or several scales, while other correlations were hypothesized to be nonsignificant. The Spearman correlation coefficient was used rather than the Pearson coefficient because not all scores followed a Gaussian distribution.

When the ratings of the two raters differed only by one unit, the mean of their ratings was assigned. When ratings differed by more than one unit, a consensual rating was assigned after discussion. These final ratings were used in this analysis and in analyses described hereunder.
Le Mer and Stoléru

Change over time
To determine whether significant changes occurred during IVF, ratings assigned to the same patients before OR and before ET were compared by the Wilcoxon matched-pairs signed-ranks test. In addition, the correlations between ratings of these two successive interviews were assessed using Spearman coefficient of correlation.

Factor analysis of scales
Only scales with a good interrater reliability (ICC > 0.60) and with a bivariate normal distribution were factor analyzed. Factors were extracted by the Principal Components method using varimax rotation (Harman, 1976). Then, the construct validity of factors was studied using the same procedure as for the individual scales.

RESULTS

Interrater Reliability
First, interrater reliability of scales 1–8 and of scale 12 was analyzed on the basis of the 65 interviews that had not been conducted by one of the two raters, that is, 41 interviews from the sample of 57 women and 24 from the sample of 29 men. Due to the later development of scales 9 and 10, the analysis of their reliability was based on 39 interviews. Moreover, as scale 11 applied only to men, the analysis was based on 24 interviews. The results are presented in Table 1. ICCs higher than 0.60 were considered to represent good agreement beyond chance, values below 0.50 poor agreement, and values between 0.50 and 0.60 fair to good agreement (Fleiss, 1981). Thus, agreement was good for 10 scales, fair to good for one scale (scale 10), and poor for one scale (scale 12). Because of the poor agreement for scale 12, its construct validity was not assessed.

Second, the same analysis performed on the ratings of interviews conducted by one of the raters yielded ICCs similar to those reported above. However, because for three scales (scales 8, 9, and 10) the ICC was below 0.50, we thought that conducting an interview may have influenced the subsequent rating on these scales. Thus, for interviews conducted by a rater, in the analyses reported hereunder, we used for these three scales the ratings of the non-interviewer.

Interviewer Effect
Two interviewers, A and B, had interviewed enough subjects (group A, n = 20; group B, n = 10) to permit a statistical comparison between ratings assigned to the groups they had interviewed. As shown in Table 1, for each scale there was no significant difference between ratings of groups A and B. Furthermore, for all the scales, the difference between the means of ratings of the two groups was less than 1 on a scale ranging from 1 to 9 (range of differences 0.10–0.97).
Assessment of Construct Validity

The matrix of correlations between the Assess_Mind scales and the other psychological measures is presented in Table 2 and Table 3, for women and men, respectively. Unless specified otherwise, results refer to measures performed on the day before OR. For each scale, we present the hypothesized correlations and the actual findings as well as statistically significant correlations that had not been hypothesized.

Scale 1. Wish to have a child
For both women and men, we had hypothesized a strong positive correlation between scale 1 and the first factor of the CPQ, which also reflected the wish for a child. This correlation was indeed significant in both genders. Because past problems might decrease the wish for a child, we had also hypothesized a nega-

---

Table 1: Interrater reliability, interviewer effect and correlation across time of ratings of Assess_Mind scales

<table>
<thead>
<tr>
<th>Assess_Mind scale</th>
<th>Interrater reliability</th>
<th>Interviewer effect</th>
<th>T1/T2 correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66</td>
<td>0.34</td>
<td>0.58**</td>
</tr>
<tr>
<td>2</td>
<td>0.69</td>
<td>0.97</td>
<td>0.71***</td>
</tr>
<tr>
<td>3</td>
<td>0.64</td>
<td>0.69</td>
<td>0.27</td>
</tr>
<tr>
<td>4</td>
<td>0.68</td>
<td>0.44</td>
<td>0.59**</td>
</tr>
<tr>
<td>5</td>
<td>0.64</td>
<td>0.81</td>
<td>0.45*</td>
</tr>
<tr>
<td>6</td>
<td>0.69</td>
<td>0.11</td>
<td>0.38</td>
</tr>
<tr>
<td>7</td>
<td>0.62</td>
<td>0.27</td>
<td>0.11</td>
</tr>
<tr>
<td>8</td>
<td>0.65</td>
<td>0.10</td>
<td>0.49*</td>
</tr>
<tr>
<td>9</td>
<td>0.77</td>
<td>0.30</td>
<td>0.76**</td>
</tr>
<tr>
<td>10</td>
<td>0.51</td>
<td>0.19</td>
<td>0.42</td>
</tr>
<tr>
<td>11</td>
<td>0.78</td>
<td>NA§</td>
<td>NA§</td>
</tr>
<tr>
<td>12</td>
<td>0.37</td>
<td>NA†</td>
<td>NA†</td>
</tr>
</tbody>
</table>

*a Scale 1: Wish to Have a Child; Sc. 2: Negative Affects Associated with the Project to Conceive a Child; Sc. 3: Negative Affects Associated with the IVF Procedure; Sc. 4: Negative Affects Associated with Infertility; Sc. 5: Negative Affects Associated with Problems not Directly Related to the Current IVF Trial; Sc. 6: Positive Affects Associated with the Project to Have a Child; Sc. 7: Vulnerability to Psychosomatic Disorganization; Sc. 8: Expression of Emotions; Sc. 9: Positive Episodic Memories Associated with the Project to Conceive a Child; Sc. 10: Negative Episodic Memories Associated with the Project to Conceive a Child; Sc. 11. Husband’s Support and Concern for Wife; Scale 12. Representational Activity Related to IVF.

*b Interrater reliability was assessed with the intraclass coefficient of correlation.

c Interviewer effect was assessed as the mean difference between the ratings assigned to two samples of interviews conducted by two different interviewers. Ratings could theoretically vary from 1 to 9.

d Spearman coefficient of correlation. T1 = Time 1, day before OR; T2 = Time 2, day of ET.

*p < 0.05; ** p < 0.01; *** p < 0.001.

§ Scale 11 applies only to men; statistics not computed because sample sizes were too small.

† Not analysed as interrater reliability was below 0.40.

Assessment of Construct Validity

The matrix of correlations between the Assess_Mind scales and the other psychological measures is presented in Table 2 and Table 3, for women and men, respectively. Unless specified otherwise, results refer to measures performed on the day before OR. For each scale, we present the hypothesized correlations and the actual findings as well as statistically significant correlations that had not been hypothesized.
### Table 2: Spearman coefficients of correlation between Assess_Mind scales and other psychological measures; women's sample.

<table>
<thead>
<tr>
<th>Non Assess_Mind Measure</th>
<th>Sc. 1</th>
<th>Sc. 2</th>
<th>Sc. 3</th>
<th>Sc. 4</th>
<th>Sc. 5</th>
<th>Sc. 6</th>
<th>Sc. 7</th>
<th>Sc. 8</th>
<th>Sc. 9</th>
<th>Sc. 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child Project Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wish for a child</td>
<td>0.52**</td>
<td>−0.04</td>
<td>−0.14</td>
<td>0.29*</td>
<td>−0.09</td>
<td>0.14</td>
<td>0.19</td>
<td>0.06</td>
<td>0.37*</td>
<td>−0.23</td>
</tr>
<tr>
<td>Marital Harmony in project to conceive</td>
<td>0.16</td>
<td>0.28</td>
<td>−0.16</td>
<td>0.00</td>
<td>−0.01</td>
<td>0.03</td>
<td>−0.01</td>
<td>0.04</td>
<td>−0.18</td>
<td>0.06</td>
</tr>
<tr>
<td>Anxieties rel. prospective parenthood</td>
<td>−0.31*</td>
<td>0.41**</td>
<td>0.35*</td>
<td>0.02</td>
<td>0.23</td>
<td>−0.35*</td>
<td>0.08</td>
<td>0.08</td>
<td>−0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>Difficulties in family history</td>
<td>−0.29*</td>
<td>−0.31*</td>
<td>0.30*</td>
<td>−0.04</td>
<td>−0.07</td>
<td>−0.30*</td>
<td>0.11</td>
<td>0.01</td>
<td>−0.43*</td>
<td>0.18</td>
</tr>
<tr>
<td>Sexual difficulties</td>
<td>−0.08</td>
<td>0.11</td>
<td>0.07</td>
<td>−0.11</td>
<td>0.12</td>
<td>−0.02</td>
<td>−0.19</td>
<td>−0.15</td>
<td>−0.20</td>
<td>0.31</td>
</tr>
<tr>
<td>Negative affects rel. infertility and IVF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative affects rel. infertility</td>
<td>0.15</td>
<td>0.17</td>
<td>0.25</td>
<td>0.32*</td>
<td>0.25</td>
<td>−0.22</td>
<td>0.24</td>
<td>0.17</td>
<td>0.14</td>
<td>0.31</td>
</tr>
<tr>
<td>Medical examinations prior to IVF</td>
<td>−0.27</td>
<td>0.20</td>
<td>0.41**</td>
<td>0.14</td>
<td>−0.05</td>
<td>−0.28*</td>
<td>0.23</td>
<td>0.08</td>
<td>−0.06</td>
<td>0.32</td>
</tr>
<tr>
<td>Treatments prior to IVF</td>
<td>−0.08</td>
<td>0.14</td>
<td>0.35*</td>
<td>0.16</td>
<td>0.12</td>
<td>−0.01</td>
<td>0.12</td>
<td>0.18</td>
<td>0.22</td>
<td>0.33</td>
</tr>
<tr>
<td>Initial steps of IVF (decision, stimulation)</td>
<td>−0.02</td>
<td>0.33*</td>
<td>0.14</td>
<td>0.04</td>
<td>0.40**</td>
<td>−0.05</td>
<td>0.15</td>
<td>0.14</td>
<td>0.02</td>
<td>−0.10</td>
</tr>
<tr>
<td>Current examinations and treatments</td>
<td>−0.18</td>
<td>0.28*</td>
<td>0.43**</td>
<td>0.17</td>
<td>0.06</td>
<td>−0.43**</td>
<td>0.30*</td>
<td>0.14</td>
<td>−0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Critical evaluation of clinic</td>
<td>−0.45**</td>
<td>0.03</td>
<td>0.25</td>
<td>0.03</td>
<td>−0.18</td>
<td>−0.42**</td>
<td>0.20</td>
<td>−0.14</td>
<td>−0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Lack of support</td>
<td>−0.39**</td>
<td>−0.09</td>
<td>−0.06</td>
<td>0.02</td>
<td>−0.16</td>
<td>−0.16</td>
<td>−0.01</td>
<td>−0.06</td>
<td>−0.29</td>
<td>−0.05</td>
</tr>
<tr>
<td>Interference with other activities</td>
<td>−0.03</td>
<td>−0.05</td>
<td>0.06</td>
<td>0.20</td>
<td>0.08</td>
<td>−0.02</td>
<td>0.29*</td>
<td>−0.03</td>
<td>−0.08</td>
<td>−0.06</td>
</tr>
<tr>
<td>Summary score of negative affect (infertility, IVF)</td>
<td>−0.21</td>
<td>0.15</td>
<td>0.42**</td>
<td>0.28*</td>
<td>0.02</td>
<td>−0.37**</td>
<td>0.31*</td>
<td>0.12</td>
<td>0.05</td>
<td>0.36*</td>
</tr>
</tbody>
</table>
Table 2: (continued)

<table>
<thead>
<tr>
<th>Non Asses_Mind Measurea</th>
<th>Sc. 1</th>
<th>Sc. 2</th>
<th>Sc. 3</th>
<th>Sc. 4</th>
<th>Sc. 5</th>
<th>Sc. 6</th>
<th>Sc. 7</th>
<th>Sc. 8</th>
<th>Sc. 9</th>
<th>Sc. 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global summary score of negative affects since beginning of current trialb</td>
<td>−0.14</td>
<td>0.27</td>
<td>0.35*</td>
<td>0.17</td>
<td>0.14</td>
<td>−0.41**</td>
<td>0.38**</td>
<td>0.19</td>
<td>−0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>State-Trait Anxiety Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Anxiety, Day 1 (before OR)c</td>
<td>−0.33*</td>
<td>0.27</td>
<td>0.41**</td>
<td>0.06</td>
<td>−0.08</td>
<td>−0.49**</td>
<td>0.27</td>
<td>0.20</td>
<td>−0.21</td>
<td>0.31</td>
</tr>
<tr>
<td>State Anxiety, Day 2 (before OR)c</td>
<td>−0.17</td>
<td>0.01</td>
<td>0.55**</td>
<td>0.44**</td>
<td>0.10</td>
<td>−0.48**</td>
<td>0.23</td>
<td>0.19</td>
<td>−0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Trait Anxiety</td>
<td>−0.12</td>
<td>0.35*</td>
<td>0.36*</td>
<td>0.19</td>
<td>0.23</td>
<td>−0.35*</td>
<td>0.30*</td>
<td>0.27</td>
<td>−0.18</td>
<td>−0.07</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>−0.09</td>
<td>0.26</td>
<td>0.34*</td>
<td>0.20</td>
<td>0.38*</td>
<td>−0.27</td>
<td>0.07</td>
<td>0.10</td>
<td>−0.31</td>
<td>0.00</td>
</tr>
<tr>
<td>Ways of Coping Checklist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem-Focused Coping</td>
<td>0.10</td>
<td>−0.28*</td>
<td>−0.37**</td>
<td>−0.08</td>
<td>−0.09</td>
<td>0.23</td>
<td>−0.13</td>
<td>−0.03</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>Wishful Thinking, Avoidance of Problem</td>
<td>0.13</td>
<td>0.07</td>
<td>−0.17</td>
<td>0.25</td>
<td>−0.01</td>
<td>−0.03</td>
<td>0.17</td>
<td>−0.14</td>
<td>0.15</td>
<td>−0.08</td>
</tr>
<tr>
<td>Seeks Social Support</td>
<td>0.08</td>
<td>−0.06</td>
<td>−0.09</td>
<td>0.02</td>
<td>0.04</td>
<td>0.12</td>
<td>−0.11</td>
<td>−0.07</td>
<td>0.15</td>
<td>−0.07</td>
</tr>
<tr>
<td>Positive Reappraisal of Stressful Situation</td>
<td>0.35*</td>
<td>−0.17</td>
<td>−0.29*</td>
<td>0.21</td>
<td>0.06</td>
<td>0.11</td>
<td>−0.12</td>
<td>0.09</td>
<td>0.01</td>
<td>−0.09</td>
</tr>
<tr>
<td>Blame Directed against Self</td>
<td>0.23</td>
<td>−0.11</td>
<td>0.18</td>
<td>0.15</td>
<td>0.21</td>
<td>0.04</td>
<td>0.18</td>
<td>0.08</td>
<td>0.07</td>
<td>−0.06</td>
</tr>
<tr>
<td>Neonatal Perception Inventory</td>
<td>−0.01</td>
<td>0.12</td>
<td>−0.02</td>
<td>0.09</td>
<td>0.12</td>
<td>0.03</td>
<td>−0.27</td>
<td>0.28</td>
<td>0.03</td>
<td>0.27</td>
</tr>
<tr>
<td>Social Desirability</td>
<td>0.20</td>
<td>−0.32*</td>
<td>−0.15</td>
<td>0.04</td>
<td>−0.13</td>
<td>0.30*</td>
<td>−0.20</td>
<td>−0.06</td>
<td>0.05</td>
<td>−0.26</td>
</tr>
</tbody>
</table>

Notes. For all scales, correlations are based on N = 57 subjects, except for scales 9 and 10 where N = 32.
a See table 1 for labels of scales and levels of probability significance. b The Global summary score of negative affects since beginning of current trial includes negative affects related, or not related, to the trial or to infertility. c Day 1 and Day 2: two days and one day before oocyte retrieval, respectively.
Table 3: Spearman coefficients of correlation between Assess_Mind scales and other psychological measures; men’s sample

<table>
<thead>
<tr>
<th>Non Assess_Mind Measure</th>
<th>Sc. 1</th>
<th>Sc. 2</th>
<th>Sc. 3</th>
<th>Sc. 4</th>
<th>Sc. 5</th>
<th>Sc. 6</th>
<th>Sc. 7</th>
<th>Sc. 8</th>
<th>Sc. 9</th>
<th>Sc. 10</th>
<th>Sc. 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child Project Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wish for a child</td>
<td>0.47*</td>
<td>−0.27</td>
<td>−0.01</td>
<td>0.09</td>
<td>−0.14</td>
<td>0.30</td>
<td>0.00</td>
<td>0.32</td>
<td>0.21</td>
<td>−0.37</td>
<td>0.12</td>
</tr>
<tr>
<td>Anxieties rel. prospective parenthood</td>
<td>−0.22</td>
<td>0.36</td>
<td>−0.04</td>
<td>0.06</td>
<td>0.27</td>
<td>−0.29</td>
<td>0.16</td>
<td>−0.15</td>
<td>−0.01</td>
<td>0.50*</td>
<td>0.12</td>
</tr>
<tr>
<td>Difficulties in family history</td>
<td>−0.36</td>
<td>0.44*</td>
<td>−0.11</td>
<td>−0.34</td>
<td>0.20</td>
<td>−0.31</td>
<td>0.16</td>
<td>−0.09</td>
<td>−0.27</td>
<td>0.35</td>
<td>−0.05</td>
</tr>
<tr>
<td>Sexual difficulties</td>
<td>−0.16</td>
<td>0.29</td>
<td>0.29</td>
<td>0.04</td>
<td>−0.05</td>
<td>−0.27</td>
<td>−0.24</td>
<td>0.04</td>
<td>−0.21</td>
<td>0.45</td>
<td>−0.23</td>
</tr>
<tr>
<td>Negative affects rel. infertility and IVF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative affects rel. infertility</td>
<td>0.09</td>
<td>−0.03</td>
<td>0.40</td>
<td>0.56**</td>
<td>0.15</td>
<td>−0.14</td>
<td>0.24</td>
<td>−0.12</td>
<td>0.02</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>Medical examinations prior to IVF</td>
<td>−0.01</td>
<td>−0.36</td>
<td>0.30</td>
<td>0.27</td>
<td>0.19</td>
<td>−0.42*</td>
<td>0.38*</td>
<td>−0.21</td>
<td>−0.19</td>
<td>0.07</td>
<td>−0.13</td>
</tr>
<tr>
<td>Treatments prior to IVF</td>
<td>−0.06</td>
<td>−0.01</td>
<td>0.24</td>
<td>0.04</td>
<td>0.33</td>
<td>−0.07</td>
<td>0.04</td>
<td>0.11</td>
<td>−0.05</td>
<td>0.42</td>
<td>0.06</td>
</tr>
<tr>
<td>Initial steps of IVF (decision, stimulation)</td>
<td>−0.32</td>
<td>−0.15</td>
<td>0.35</td>
<td>0.43*</td>
<td>0.45*</td>
<td>−0.36</td>
<td>0.17</td>
<td>−0.08</td>
<td>0.12</td>
<td>0.48</td>
<td>0.22</td>
</tr>
<tr>
<td>Current examinations and treatments</td>
<td>−0.24</td>
<td>−0.28</td>
<td>0.07</td>
<td>0.05</td>
<td>0.51**</td>
<td>−0.31</td>
<td>0.25</td>
<td>0.05</td>
<td>−0.29</td>
<td>0.25</td>
<td>−0.01</td>
</tr>
<tr>
<td>Critical evaluation of clinic</td>
<td>−0.33</td>
<td>−0.07</td>
<td>0.23</td>
<td>0.46*</td>
<td>0.30</td>
<td>−0.12</td>
<td>−0.13</td>
<td>−0.10</td>
<td>−0.15</td>
<td>0.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Lack of support</td>
<td>0.05</td>
<td>−0.14</td>
<td>−0.46*</td>
<td>−0.01</td>
<td>−0.03</td>
<td>−0.24</td>
<td>0.16</td>
<td>−0.30</td>
<td>0.25</td>
<td>−0.27</td>
<td>−0.30</td>
</tr>
<tr>
<td>Interference with other activities</td>
<td>−0.05</td>
<td>0.19</td>
<td>0.16</td>
<td>0.33</td>
<td>0.29</td>
<td>−0.20</td>
<td>0.10</td>
<td>−0.20</td>
<td>0.01</td>
<td>0.43</td>
<td>0.14</td>
</tr>
<tr>
<td>Summary score of negative affects (infertility, IVF)</td>
<td>−0.11</td>
<td>−0.03</td>
<td>0.24</td>
<td>0.49**</td>
<td>0.42*</td>
<td>−0.40*</td>
<td>0.33</td>
<td>−0.21</td>
<td>−0.20</td>
<td>0.43</td>
<td>−0.03</td>
</tr>
</tbody>
</table>
Table 3: (continued)

<table>
<thead>
<tr>
<th>Non Assess_Mind Measurea</th>
<th>Sc. 1</th>
<th>Sc. 2</th>
<th>Sc. 3</th>
<th>Sc. 4</th>
<th>Sc. 5</th>
<th>Sc. 6</th>
<th>Sc. 7</th>
<th>Sc. 8</th>
<th>Sc. 9</th>
<th>Sc. 10</th>
<th>Sc. 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global summary score of negative affects since beginning of current trialb</td>
<td>−0.16</td>
<td>0.06</td>
<td>0.36</td>
<td>0.45*</td>
<td>0.45*</td>
<td>−0.35</td>
<td>0.25</td>
<td>−0.13</td>
<td>−0.49*</td>
<td>0.40</td>
<td>0.01</td>
</tr>
<tr>
<td>State-Trait Anxiety Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Anxiety, Day 1c (before OR)</td>
<td>−0.19</td>
<td>−0.02</td>
<td>0.08</td>
<td>0.45*</td>
<td>0.34</td>
<td>−0.44*</td>
<td>0.50*</td>
<td>0.03</td>
<td>−0.28</td>
<td>0.34</td>
<td>−0.07</td>
</tr>
<tr>
<td>State Anxiety, Day 2c (before OR)</td>
<td>−0.35</td>
<td>−0.09</td>
<td>0.42*</td>
<td>0.33</td>
<td>0.30</td>
<td>−0.42*</td>
<td>0.39</td>
<td>−0.20</td>
<td>−0.52*</td>
<td>0.06</td>
<td>−0.06</td>
</tr>
<tr>
<td>Trait Anxiety</td>
<td>−0.33</td>
<td>0.11</td>
<td>0.32</td>
<td>0.46*</td>
<td>0.32</td>
<td>−0.47*</td>
<td>0.35</td>
<td>−0.06</td>
<td>−0.29</td>
<td>0.24</td>
<td>−0.02</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>−0.26</td>
<td>0.11</td>
<td>0.33</td>
<td>0.52*</td>
<td>0.52*</td>
<td>−0.16</td>
<td>0.24</td>
<td>0.00</td>
<td>−0.20</td>
<td>0.48*</td>
<td>0.19</td>
</tr>
<tr>
<td>Ways of Coping Checklist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem-Focused Coping</td>
<td>0.02</td>
<td>−0.17</td>
<td>−0.10</td>
<td>−0.31</td>
<td>0.16</td>
<td>−0.08</td>
<td>−0.24</td>
<td>−0.16</td>
<td>0.14</td>
<td>−0.29</td>
<td>−0.17</td>
</tr>
<tr>
<td>Wishful Thinking, Avoidance of Problem</td>
<td>−0.23</td>
<td>0.13</td>
<td>0.03</td>
<td>0.23</td>
<td>0.37</td>
<td>−0.31</td>
<td>0.06</td>
<td>−0.11</td>
<td>−0.09</td>
<td>0.40</td>
<td>−0.08</td>
</tr>
<tr>
<td>Seeks Social Support</td>
<td>0.00</td>
<td>−0.03</td>
<td>0.33</td>
<td>−0.09</td>
<td>0.35</td>
<td>0.18</td>
<td>0.04</td>
<td>−0.04</td>
<td>−0.02</td>
<td>−0.37</td>
<td>0.21</td>
</tr>
<tr>
<td>Positive Reappraisal of Stressful Situation</td>
<td>0.02</td>
<td>−0.23</td>
<td>0.15</td>
<td>0.00</td>
<td>0.41*</td>
<td>−0.13</td>
<td>0.02</td>
<td>0.11</td>
<td>0.21</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Blame Directed against Self</td>
<td>−0.54**</td>
<td>0.09</td>
<td>−0.02</td>
<td>0.21</td>
<td>0.27</td>
<td>−0.48**</td>
<td>0.11</td>
<td>−0.22</td>
<td>−0.36</td>
<td>0.01</td>
<td>−0.23</td>
</tr>
<tr>
<td>Neonatal Perception Inventory</td>
<td>0.15</td>
<td>0.35</td>
<td>0.13</td>
<td>0.26</td>
<td>0.09</td>
<td>0.06</td>
<td>−0.10</td>
<td>0.29</td>
<td>0.20</td>
<td>0.29</td>
<td>0.45</td>
</tr>
<tr>
<td>Social Desirability</td>
<td>0.33</td>
<td>−0.09</td>
<td>−0.19</td>
<td>−0.56**</td>
<td>−0.23</td>
<td>0.26</td>
<td>−0.23</td>
<td>−0.04</td>
<td>0.26</td>
<td>−0.34</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Notes: For all scales, correlations are based on N = 29 subjects, except for scales 9 and 10 where N = 18.
a See table I for labels of scales and levels of probability significance. b The Global summary score of negative affects related, or not related, to the trial or to infertility. c Day 1 and Day 2: two days and one day before oocyte retrieval, respectively.
tive correlation between scale 1 and the CPQ score referring to problems in family history related to procreation. This correlation was significant in women.

Moreover, in women, scale 1 was negatively correlated with several scores, that is, anxieties related to prospective parenthood, level of criticism of current medical care and clinic, perceived lack of support from others, and state anxiety measured two days before OR. Finally, in women, scale 1 was positively correlated with the fourth factor of the *Ways of Coping Checklist*, that is, positive reappraisal of stressful situations, whereas in men it was negatively correlated with the fifth factor, that is, blame directed against self.

Scale 2. Negative affects associated with the project to conceive a child
A classical psychoanalytic hypothesis about the etiology of psychogenic infertility is that the apparent wish to have a child may hide various anxieties related to having a child and/or becoming a parent (Deutsch, 1945; Benedek, 1952; Bos and Cleghorn, 1958). On the basis of this hypothesis, and as we were not aware of any instrument designed to measure this kind of anxiety, when developing the CPQ we included six questions to assess various anxieties associated with the potential success of the project to have a child. We then constructed a score of anxieties associated with prospective parenthood (AAPP) by summing responses to these six items. We hypothesized that scale 2 and the AAPP score would be correlated. Actually, this correlation was significant in women and marginally significant in men ($p = 0.05$). We had also hypothesized a positive correlation between scale 2 and the NPI, which was not confirmed.

As hypothesized, we found a negative correlation between scale 2 and the social desirability measure, but only in women. In addition, in women, scale 2 was positively correlated with the score assessing how stressful the initial steps of the current trial had been and with trait anxiety. In men, scale 2 was negatively correlated with two individual items of the CPQ, that is, how frequently subject imagined himself as a father ($R_s = -0.37; p < 0.05$) and how intense was subject’s desire for a child ($R_s = -0.45; p < 0.05$). Conversely, in men, scale 2 was positively correlated with the score of difficulties in family history related to procreation. In women, surprisingly, there was a negative correlation between scale 2 and the score of difficulties in family history related to procreation. To understand this finding, it should be noted that a great number of women ($n = 27; 47.4\%$) did not express any negative affects associated with the project to conceive a child (score on scale 2 $= 1$). It is likely to be difficult for a patient treated by IVF to express any such affects, even to a researcher not involved in the treatment. Indeed, within the subsample of women ($n = 28$) who expressed such affects (score $>1$ on scale 2) there was a positive correlation between scale 2 and the score of difficulties in family history related to procreation ($R_s = 0.38; p < 0.05$). Similarly, in men scoring higher than 1 on scale 2 ($n = 14$), this correlation was significant ($R_s = 0.62; p < 0.05$).
Scale 3. Negative affects associated with the IVF procedure
This scale was designed to measure patients’ anxiety in response to medical procedures of the IVF trial (Boivin and Takefman, 1996). Because state anxiety as measured by the STAI was likely to reflect multiple sources of anxiety, including anxieties triggered by impending OR, we hypothesized a moderately high positive correlation between scale 3 and state anxiety measured on the day before OR. This correlation was significant in both genders. We also predicted that scale 3 would be positively correlated with trait anxiety. However, as trait anxiety was much less susceptible to day-to-day fluctuations related to external acute stressors, we hypothesized that this correlation would be rather low. The observed correlation was significant only in women.

We hypothesized a significantly positive correlation between scale 3 and the score on the PSS, but only in the women’s sample, more directly concerned with the medical procedures. This correlation was indeed significant only in women. As previous negative IVF experiences were likely to influence current feelings, we predicted, again only in the women’s sample, a positive correlation between scale 3 and negative affects associated with medical examinations preceding the current IVF trial, as well as between scale 3 and negative affects associated with treatment procedures preceding the current IVF trial. Both correlations were significant. Interestingly, and as predicted, in women there was a negative correlation between scale 3 and the first dimension of coping, that is, problem-focused coping. Finally, in men only, there was a paradoxically negative correlation between scale 3 and the perceived lack of support during the IVF procedure.

Scale 4. Negative affects associated with infertility
Scale 4 was developed as a measure of the distress induced by infertility and by the related fear that the current trial might fail. We predicted a positive correlation between scale 4 and the measure of wish for a child derived from the CPQ. The rationale for this prediction was that the desire to have a child was a prerequisite for suffering from being infertile and for fearing a failure of IVF. The observed correlation was significant in women. As hypothesized, we found in both genders a positive correlation between scale 4 and the score of negative affects associated with infertility.

Interestingly, the pattern of correlations of scale 4 with the other scores assessing the experience of the IVF trial was different in women and men. While in men the summary score of negative affect associated with current trial (affective responses to IVF and to infertility) was strongly correlated with scale 4 but not with scale 3, in women the opposite pattern was observed. Similarly, in women the global summary score of negative affects experienced since beginning of current trial (whether these affects were related or not, to the trial) was correlated with Scale 3, that is, with the scale of negative affects associated with the IVF procedures, but not with Scale 4. The reverse was true for men. These
diverging patterns suggest that for women negative feelings associated with IVF stemmed mainly from medical procedures rather than from the fear of failure, whereas for men the reverse was true. This is in keeping with the very different manners in which women and men were involved in medical procedures. Similarly, in men only, scale 4 was correlated with other scores related to the experience of the current IVF trial.

- Scale 4 was correlated with the score of criticism of medical care and clinic. Thus, the more a man experienced negative feelings in response to infertility, including the fear of failure of current trial, the more he assessed the clinical management as poor.
- Scale 4 was positively correlated with the score of negative affects associated with the initial steps of IVF.

We predicted a positive correlation between scale 4 and state anxiety measured on the same day. The observed correlation was significant in women but not in men. In men, it is the level of state anxiety measured one day before the Assess_Mind that was correlated with scale 4. If the correlation observed in women truly reflected that state anxiety was partly based on the fear of failure of current trial, then this correlation should decrease sharply in women after successful fertilization. Indeed, in women, after fertilization was known to be successful, scale 4 and state anxiety were no longer significantly correlated (not shown on tables). We also hypothesized that scale 4 would be positively correlated with trait anxiety and with the PSS. These correlations were significant in men. Again, in men, scale 4 was negatively correlated with social desirability. Therefore, the above analysis of scale 4 correlations was performed again controlling for social desirability. All significant correlations were confirmed, except the correlation with the PSS.

Scale 5. Negative affects associated with problems not directly related to the current IVF trial
Scale 5 was designed to assess negative affects with no apparent relation with procreation and/or IVF. We hypothesized a positive correlation with the state anxiety scale, the trait anxiety scale, and with the PSS, and a negative correlation with the first factor of the WCC, that is, problem-oriented coping. In women, we found no significant correlation between scale 5 and state anxiety. In men, the correlation between scale 5 and state anxiety measured two days after OR (not shown in Table 3) was $R_s = 0.46$ ($p < 0.05$). Thus, whereas in women, state anxiety was not related to sources of anxieties other than the project to conceive a child, in men two days after OR state anxiety was related with these sources of anxiety. We found a positive correlation between scale 5 and the PSS in both genders. Only in men, a significant positive correlation was found between scale 5 and a factor of the WCC, that is, positive re-appraisal of stressful situation.
Interestingly, and as predicted, in women there were very few correlations between scale 5 and variables measuring the emotional impact of infertility and/or IVF medical procedures. The only significant correlation was between scale 5 and the score of negative affects associated with initial steps of IVF. Scale 5 included the emotional problems encountered by women in trying both to continue to work and to follow time-consuming medical procedures, often without explaining this problem to their hierarchy. Therefore, early stages of the IVF procedure were likely to have contributed to professional stress. Reciprocally, negative emotions unrelated to IVF, but experienced at the time of the initial steps of IVF, were likely to have aggravated the stress related to these procedures. In men, by contrast, scale 5 was related to several measures of negative affects related to IVF. Thus, the higher the men's difficulties in areas unrelated to IVF, and the higher their difficulty to cope with IVF treatment.

Scale 6. Positive affects associated with the project to have a child
We hypothesized that scale 6 would be correlated: (i) positively with factor I of the CPQ, that is, wish to have a child; and negatively with (ii) the various scores of negative affects associated with the current IVF trial; (iii) the PSS; (iv) state anxiety measured before the gynecologist’s feedback about the outcome of the fertilization step; and (v) the NPI. Actually, in women the CPQ score of AAPP was negatively correlated with scale 6, and there was a negative correlation between scale 6 and several scores of negative affects associated with examinations and treatments performed during current IVF trial. In men, scale 6 was negatively correlated with the score of negative affects associated with medical examinations performed before the current IVF trial and with the summary score of negative affects associated with current trial (infertility, IVF). As hypothesized, in both genders, scale 6 was negatively correlated with state anxiety and with trait anxiety. In women, there was a positive correlation between scale 6 and social desirability. After the correlational analysis was performed again, controlling for social desirability, all above-mentioned correlations remained significant, except the correlation with the score of anxiety associated with prospective parenthood.

Scale 7. Vulnerability to psychosomatic disorganization
The rationale for developing this scale was that various psychosomatic disturbances in response to anxiety-provoking situations are more frequent or severe in infertile than in fertile women (Kemeter, 1988). The rating of this scale was based on recording past and current psychosomatic disorders or episodes mentioned during the Assess_Mind interview. We hypothesized a correlation between scale 7 and state anxiety. The actual correlation with state anxiety measured two days before OR was significant in men, but not in women. However, in women scale 7 was correlated with trait anxiety. In women, scale 7 was also significantly correlated with several variables related to the experience of IVF.
Scale 8. Expression of emotions
As scale 8 was designed to measure how intensely patients felt and expressed their emotions, we hypothesized that it would be correlated with scores derived from the questionnaire investigating the emotional impact of the IVF procedure. However, the observed correlations were not significant.

Scales 9 and 10. Episodic memories associated with the project to conceive a child
We developed two scales related to memories: scale 9, Positive episodic memories, that is, memories denoting positive attitudes towards the project to conceive, and scale 10, Negative episodic memories, denoting negative attitudes towards this project. We hypothesized that these two scales would be correlated with state anxiety, with a negative correlation for scale 9 and a positive one for scale 10. In men, but not in women, scale 9 was indeed negatively correlated with state anxiety measured on the same day.

In addition, as hypothesized, in women scale 9 was negatively correlated with the score of difficulties in family history related to procreation. In women, scale 9 was also positively correlated with wish for a child, that is, factor I of the CPQ. In men, scale 9 was negatively correlated with the global summary score of negative affects experienced since beginning of current trial, whether these affects were related, or not, to the trial.

In women, there was a correlation between scale 10 and the summary score of negative affect associated with current trial (infertility, IVF). In men, scale 10 was positively correlated with the score of anxieties associated with prospective parenthood and with the score on the PSS.

Scale 11. Husband’s support and concern for wife
This scale was rated only on transcripts of men’s interviews. It was not significantly correlated with psychological measures derived from other instruments.

As shown in Table 2 and Table 3, most of the correlation coefficients between social desirability and Assess_Mind scales were not significant. The median value of the coefficients of correlation was 0.15 in women and 0.23 in men (when calculating this value, only the absolute value of coefficients was considered). These values are slightly lower than the medians of the coefficients of correlation between social desirability scores and psychological variables derived from instruments other than the Assess_Mind, that is, 0.18 in women and 0.32 in men. In other words, social desirability did not affect Assess_Mind ratings more than it affected other psychological measures.

Change Over Time

In 21 women (36.8%), at least one embryo was obtained and a second Assess_Mind interview was conducted before ET. Ratings assigned to the same women
on the two interviews were not statistically different, except for scale 6, positive affects associated with the project to have a child, where ratings increased (means (±standard deviation [SD]) = 4.7 (±1.3) on time 1 and 5.6 (±1.7) on time 2; Wilcoxon matched-pairs signed-ranks test \( Z = -2.13; p < 0.05 \)). Correlations between ratings of interviews performed before OR and before ET varied widely among scales (Table 1). Coefficients of correlation were low for scale 3, negative affect associated with IVF procedure, and scale 7, vulnerability to psychosomatic disorganization. For these scales, the ratings before OR were probably influenced by medical procedures unique to OR (surgical intervention, hormonal treatment), which is likely to account for the low correlation between OR and ET ratings. The corresponding analyses were not performed in men because the sample sizes were too small.

**Factor Analysis of Assess_Mind Scales**

Factor analysis involved all the scales, except scales 2 and 7 because of their nongaussian distribution, and scale 12 because of insufficient interrater reliability. Factor analysis was performed only in women because the size of the men's sample was too small. Factor analysis yielded four factors, explaining 73.9% of the variance of ratings. All the scales loaded strongly on one of the four factors (loading ≥ 0.65), except scale 8, expression of emotions, which loaded on three factors, which is not surprising given its meaning. The first factor, which was labeled Desire to Have a Child, was highly correlated with scales 1, Wish to have a child, scale 4, Negative affects associated with infertility, scale 6, Positive affects associated with the project to have a child, and scale 9, Positive episodic memories associated with the project to have a child. Factor II was mainly related to scale 3, Negative affects associated with IVF procedure. Factor III reflected almost exclusively scale 5, Negative affects not directly related to IVF trial. Finally, factor IV was essentially related to scale 10, Negative episodic memories associated with project to conceive a child.

Correlations of these factors with measures derived from other instruments are presented in Table 4. Factor I was positively correlated with CPQ factor I, which was also a measure of the desire to have a child, and negatively correlated to state anxiety measured two days before OR. Factor II, Negative affects related to IVF procedure, was positively correlated with state anxiety measured one day before OR. As expected, factor II was also related to several measures of the negative affective responses to the IVF procedure. Factor III, Negative affects not directly related to IVF trial, was correlated with trait anxiety, perceived stress, Negative affects associated with initial steps of IVF, and the global summary score of negative affects experienced since beginning of current trial. Factor IV, Negative episodic memories related to project to conceive, was negatively correlated with CPQ factor I, Wish to Have a Child, and positively correlated with several measures of the negative affective responses to the IVF procedure. Finally, none of these factors was correlated with social desirability.
DISCUSSION

For 11 of the 12 scales rated by non-interviewers, interrater reliability was fair to excellent (median ICC = 0.66; range = 0.51–0.78) (Fleiss, 1981). This is a
notable result for measures derived from an interview based on open questions focusing on inner experience rather than on factual information or on symptoms. These values are of the same order of magnitude as those obtained for commonly used psychiatric rating scales focusing on symptoms (e.g., Montgomery and Asberg, 1979). In addition, these results are similar or compare favorably with those reported for rating scales derived from other semistructured interviews comparable to the Assess_Mind. Thus, as regards scales applied to semistructured interviews to specify the psychodynamic profile of patients, intrarater reliability varied greatly between scales (0.15 to 0.61) (Aguilar et al., 1996) or was mostly good to excellent (Weinryb and Rössel, 1991). In a psychoanalytically based study of subjects’ interpersonal relations based on the Personal Relatedness Profile, Hobson et al. (1998) assessed intrarater reliability for judgments of characteristics of interpersonal relatedness: using Kendall’s $W$ as a measure of agreement among judges, the median reliability of items was $W = 0.34$ (range 0.21–0.60). Using a modification of the Personal Relatedness Profile to assess couple relationships, Lanman et al. (2003) also reported good intrarater reliability, with a median ICC of 0.77 (range 0.38–0.94).

In the study of defense mechanisms through interviews, Vaillant et al. (1986) obtained good intrarater reliability. However, in other studies of defense mechanisms, ICCs ranged from 0.30 to 0.50 for individual scales (Busch et al., 1995), or the median ICC was about 0.61, with about 50% of ICCs falling below 0.60 (Jacobson et al., 1986; Bloch et al., 1993; Perry, 2001), that is, below a commonly used threshold for good to excellent ICC. Intrarater reliability of scales applied to an interview to evaluate the suitability of patients for psychodynamic psychotherapy ranged from 0.68 to 0.80 (Rosenbaum et al., 1997). In a study of motivational interviewing, ICCs for estimating the reliability of characteristics of patients and therapists ranged from 0.51 to 0.80 (Moyers et al., 2003). Finally, good intrarater reliability has been achieved for rating scales applied to the study of adult attachment patterns (Main and Goldwyn, 1984). In the latter instance and in the Assess_Mind interview, good intrarater reliability was probably partly due to the fact that rating was performed on the basis of transcripts. Overall, the feasibility of achieving good intrarater reliability in rating semistructured interviews needs to be underlined, as it is frequently overlooked. Overlooking these findings leads to wrongly assume that structured interviews are the only method of achieving intrarater reliability.

Results of the analysis of the ratings of interviews conducted by one of the raters showed that for three scales the ICC fell below 0.50. Thus, having conducted the interview might have influenced subsequent ratings on these scales. This effect may be related to the enduring clinical impression, including the affective reactions, felt by the rater who had been the interviewer. There was no significant interviewer effect for any of the rating scales. In other words, for all scales the means of ratings assigned to interviews conducted by interviewers A and B did not differ significantly. Although the power of the test to demonstrate the latter finding was low, the absolute difference between ratings of these two
subsamples of interviews was very small. This result might be related to: (i) the
detailed written rules that must be followed to administer the Assess_Mind;
(ii) the thorough training of interviewers; and (iii) the close ongoing supervision
of interviews, as the audiotape of about one in three interviews was studied by
the coordinator of the study and discussed with the interviewers. However,
because only ratings from two interviewers could be used to assess the inter-
viewer effect, the obtained result needs to be replicated with other interviewers.
In addition, to conduct a formal interinterviewer reliability study, each patient
should have been interviewed by two interviewers with a short in-between
interval. Within the present study, this was hardly feasible: (i) because the psy-
chological state of patients changed rapidly, even over a one-day period, in
response to the IVF-procedures and (ii) because a second interview, close to the
first, would have made this study too demanding for patients already submitted
to numerous medical procedures. Nonetheless, the results mentioned above
suggest that such interinterviewer reliability could be attained with this instru-
ment. This would be important, as one major criticism addressed to semistruc-
tured interviews is their alleged lack of interinterviewer reliability (Patton,
1980).

As the Assess_Mind was designed to evaluate the patients’ current state of
mind and as the psychological state of patients changed rapidly, it was not
meaningful to perform a study of the test–retest reliability of ratings in the sub-
sample interviewed on two occasions. Depending on their specific content, only
some scales were likely to be affected by events occurring between the two
interviews, which probably explains why correlations between ratings on Time
1 and Time 2 varied widely among scales. The content of two scales, negative
affects associated with the IVF procedure, and vulnerability to psychosomatic
disorganization, was directly related to events occurring between the two inter-
views (relief from anxiety related to impending surgery, cessation of hormonal
stimulation), which is likely to account for the low correlation found for those
scales between times 1 and 2. By contrast, for wish for a child, and negative
affects associated with the project to conceive a child, events between interviews
were less likely to affect ratings and correlations were much higher.

Regarding the comparison of Time 1 and Time 2 ratings, although a test–
retest reliability study of interviews separated by a short interval remains to be
performed in a future study where less change would be expected between inter-
views, the variations of ratings from Time 1 to Time 2 may reflect actual change
over time rather than lack of reliability. Thus, the significant increase of ratings
on scale 6, positive affects associated with the project to have a child, is likely
to be related to a major change of the medical situation between times 1 and 2:
whereas before OR patients tended to be anxious about impending surgery,
before ET they felt relieved from this anxiety and that success had been achieved
at least for the fertilization step (Merari et al., 1992). While the other scales did
not vary significantly, the power of the test was low because of the small size of
the subsample with successful fertilization (n = 21).
Regarding construct validity, most scales were correlated in a meaningful way with various questionnaire measures. Exceptions were scale 8, expression of emotions, and scale 11, husband’s support and concern for wife. For all other scales, the majority of hypotheses about their correlations with conceptually related measures derived from other instruments was confirmed. Within the set of correlations that had been hypothesized between Assess_Mind scales and other psychological measures, the median value of the coefficients of correlation was 0.32 in women and 0.42 in men. While these values fall in the moderate range of magnitude, it should be stressed that in many cases the variables whose correlations were tested were not quite identical conceptually. For instance, scale 6, positive affects associated with the project to have a child, was not conceptually identical with factor I of the Child Project Questionnaire, which was essentially a measure of the intensity of the wish to have a child. When only the correlations between conceptually very close measures were considered, for example scale 1 and factor I of the CPQ, the correlations were much higher (medians 0.45 in women, 0.47 in men).

Moreover, for some of the constructed scales, that is, scales 2, 6, 7, 9, 10, and 11, we were not aware of existing scales against which to evaluate their construct validity. However, one limitation of the study was not to use one of the existing measures of the expression of emotion (e.g. Lane et al., 1990). Another limitation was that we were not able to make sure that the STAI and PSS questionnaires were completed at home on the appropriate days. However, some aspects of results support the view that subjects did complete the questionnaires appropriately. Thus, after fertilization was known to be successful, women’s scores on scale 4 (negative affects associated with infertility) and state anxiety were no longer significantly correlated.

Finding high coefficients of correlation between Assess_Mind scales and psychological measures derived from other instruments is a mixed blessing: on the one hand, it shows that the instrument being developed does measure the same constructs as previously validated instruments, a classical evidence for construct validity; on the other, very high coefficients of correlation would lead to ask what is the unique contribution of the novel instrument. Similarly, if using a novel instrument is more time-consuming than already existing instruments, potential users are right to expect that the newly developed measures be related to – but distinct from – those derived from existing instruments. In the absence of “gold standards” in psychopathology, a future comparative study of the respective predictive validities of the Assess_Mind scales and of other psychological measurements could be helpful. For instance, both kinds of measurements could be used as predictors of the outcome of IVF trials and compared on their predictive, that is, criterion-related, validity. A higher predictive validity for the new scales would support their use rather than using former similar scales.

In validating psychological measures, it is important not only to demonstrate significant correlations with similar constructs, but also to show the lack of sig-
significant correlations with theoretically unrelated measures (Kerlinger, 1986). It is therefore important that most of the scales were not significantly correlated with social desirability.

The 11 scales were initially conceived as distinct measures, which led us to present separately the psychometric characteristics of each of them, instead of restricting the presentation to the factor analysis. In addition, given the relative complexity of individual scales, this allowed a deeper appreciation of the meaning of obtained factors, which are themselves composed of these scales.

The psychometric characteristics of the Assess_Mind suggest it could be useful in clinical settings such as infertility clinics. Furthermore, at least for nine scales interrater reliability suggests that ratings may be performed by the interviewer, which would make the procedure practicable in a clinical setting.

Finally, in this paper we have considered the validity of the Assess_Mind within the domain of reproductive psychology. Further studies need to be conducted to demonstrate its reliability and its validity in other areas. This implies to conduct Assess_Mind interviews in samples of patients presenting other specific psychopathological problems and to develop corresponding scales. Such a validation study of the Assess_Mind in the assessment of major depressive episodes has been initiated and pilot studies are underway in child psychiatry and geriatric psychiatry.

ACKNOWLEDGEMENTS

We thank Organon Laboratory for financial support, D Cornet MD for the referral of patients, S Zerah PhD for facilities provided at the IVF clinic, I Assouline and S Bimbard for careful preparation of transcripts and of the manuscript, and B Le Roux, I Celestin, S Honlet, V Bollinches, and V Plard who administered interviews.

REFERENCES


Reliability and validity of assessing the mind


Marie-Noëlle Le Mer and Serge Stoléru
Unité 742 Inserm
Université Pierre et Marie Curie
Boîte 23
9 quai Saint Bernard
75252 Paris Cedex 05
France
(serge.stoleru@snv.jussieu.fr)