A self-rating scale of distress

ROBERT KELLNER* AND BRIAN F. SHEFFIELD°

From the Department of Psychiatry, University of Liverpool

SYNOPSIS Results of studies with various versions of the Symptom Rating Test are described. The test was designed to measure changes in the symptoms of neurotic patients participating in experiments in therapeutics such as drug trials. In all studies the test scores discriminated significantly between psychiatric patients and normals. In drug trials the test was found to be effective in discriminating between the responses to psychotropic drugs and to placebo. The findings suggest that the test is a valid and reliable measure of distress.

In the construction of a questionnaire, a clear distinction should be made between the measurement of a trait (a long-standing disposition) and distress (a temporary and changeable state). Responses to items in a personality inventory should be stable over time, but responses to items in a distress scale should change over time and measure changes in the clinical state of the patient (Kellner, 1967a).

The purpose of the present paper is to describe a self-rating test, the Symptom Rating Test (SRT), which was designed to measure changes in distress in experiments in therapeutics—for example, in drug trials. There are several forms of the SRT. Most studies have been carried out with the Week form, in which the patient rates symptoms he experienced during the past week, and the Day form, in which he rates the symptoms he experienced on the day of testing.

The studies have been carried out either with test cards (TC) or with the pencil and paper version (PP). A few studies were carried out with the short paper and pencil version and these are summarized separately. Except where indicated, the Spearman Rank Order Correlation Coefficient was used (Siegel, 1956). Results at or below the conventional 5% level of probability were regarded as statistically significant.

The SRT consists of a check list of 38 symptoms, 15 are somatic and 23 psychological, and the patient rates each symptom he has marked on the check list on a four-point self-rating scale. The short version contains self-rating scales only.

DEVELOPMENT OF TEST

Each stage of the development of the test was based on research findings (Kellner, 1967a).

The check list was compiled from the complaints of 100 consecutive neurotic patients. The aim was to compile a short, but comprehensive, check list of symptoms, using expressions easily understood by most patients. The check list contains 38 items, but covers a wider range of symptoms; this was achieved in part by constructing items consisting of more than one similar complaint. There is a space available for idiosyncratic symptoms which are not included in the check list. The check list appears to contain most common neurotic symptoms.

The test card version of the SRT consists of a semi-structured interview, based on the check list, during which Target Symptoms are selected. Lipman, Cole, Park, and Rickels (1965) found that the score of Target Symptoms discriminated more effectively between an antianxiety drug and placebo than the total score of a distress scale. Each Target Symptom is self-rated by the patient with randomized test cards. The cues which are typed on test cards were chosen from expressions used by the patients. Initially the patient rated each symptom on three dimensions intensity, frequency, and duration. Many scales with different kinds of cues and different num-

*Present address: Department of Psychiatry, School of Medicine, University of New Mexico, Albuquerque, N. M.
°The Department of Psychology, Wesham Park Hospital, Kirkham, Lancs.
bers of cues were tried. Only those cues were retained which literate neurotic patients could rank without difficulty. Scales which patients found difficult to use were discarded. The remaining scales were validated in two ways: they were administered to neurotic patients and normals, and to neurotic patients before and after treatment. The scales which (1) discriminated between neurotics and normals, and (2) changed in neurotic patients after treatment in the expected direction, were regarded as valid and sensitive to change (Sheffield and Kellner, 1970). These were labelled the ‘Main Scales’. The Main Scales are scales of intensity of symptoms and frequency of occurrence. The use of Main Scales alone was found to be as effective as using three scales for each symptom in discriminating between neurotics and normals and about as effective as using three scales in discriminating between the effects of drugs and placebo in two drug trials (Kellner, 1967a; Kellner, Kelly, and Sheffield, 1968; Kellner and Sheffield, 1969). The scales were shortened into four-point scales by eliminating cues which were rarely used by patients.

Positional set was found to play only a small part in the self-rating of neurotic symptoms on a short graphic scale (Bedford and Edington, 1968). The test was, therefore, converted into a more economic pencil and paper version.

There is a difference of opinion whether a continuous line or a broken line should be used for self-rating scales (Guilford, 1954). Battle, Imber, Hoehn-Saric, Stone, Nash, and Frank (1966) found that more than half the patients preferred a ‘box-scale’—that is, a broken line to a continuous line. In the SRT, the cues are concentrated at points by partial staking of the words in columns and patients are asked to draw a circle around the statement which describes their condition; thus, they are instructed to treat the scale as a broken line.

The construction of three subscales was based on several published factor analyses of symptoms. The fourth (inadequacy) subscale is not a single factor and consists mainly of feelings of inferiority and cognitive symptoms. The subscales are listed in Table 5.

The analysis of the results of several studies showed that there was no advantage in increasing the number of items beyond a certain point. A short version was constructed in which the number of items was reduced to 30. In the studies which are summarized below, the results obtained with the check list were similar to those obtained with the self-rating scales; therefore, in the short version the check list was omitted. No significant differences were found between the number of patients preferring the original self-rating scale and those preferring a box scale, and the simpler box scale was adopted (Kellner, 1972b). Since the differences obtained with various Main Scales were small, one Main Scale is used throughout instead of different scales for different symptoms; the cues of this scale are similar to those used in several self-rating scales of symptoms (Clyde, 1963; McNair and Lorr, 1964; Lipman, Covi, Rickels, Uhlenhuth, and Lazar, 1968).

For the purpose of comparison, in several studies described below, the SRT was administered together with three self-rating tests which have been extensively validated and have been shown to be valid measures of personality: the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953), the Maudsley Personality Inventory (MPI) (Eysenck, 1959), and the similar Eysenck Personality Inventory (EPI) (Eysenck and Eysenck, 1964).

**RELIABILITY (WEEK FORM)**

The test-retest reliability of a personality inventory is evidence for its temporal stability; a test which aims at measuring stable personality traits should have largely unaltered scores with the passage of time. Since the SRT appears to measure distress—a changeable state-retest after a long time interval would measure an altered state in a large proportion of patients. The function of the SRT is analogous to that of a clinical thermometer rather than to that of a tape measure.

Neurotic outpatients (N28) were readministered the Week form (test cards) after an interval of 24 hours; the test-retest correlation was +0.94. The study was repeated with inpatients (N40) with the pencil and paper version and the test-retest correlation was +0.92. The MAS was administered together with the pencil and paper version for the purpose of comparison.
the test-retest correlation of the MAS was +0.82.

The conventional split half reliability was not carried out because the items are not psychometrically equivalent. The split half reliability of changes of SRT scores (TC) in neurotic outpatients (N40) after one month was +0.89.

Four descriptive cues for the rating of each symptom were found to be adequate; increasing the number of cues seemed to decrease the reliability of the scale.

VALIDITY OF THE CHECK LIST (WEEK FORM)
The check list was validated by administering it to neurotic outpatients (N40) and normal subjects (N43) as a semistructured interview and to neurotic patients (outpatients, inpatients, and day-patients) and normal subjects in six separate studies as the pencil and paper version of the SRT. In all studies the check list score discriminated significantly between neurotics and normals (P<0.001) (Kellner, 1967a; Kellner and Sheffield, 1967; Morton, 1968; Kellner, 1972b).

The SRT check list (PP) was given to 100 neurotic patients (largely inpatients and day-patients) and 100 normals. Thirty-seven items discriminated between the two groups with P<0.001, and two items did not discriminate significantly. This study was carried out in England and repeated in New Mexico. In the second study all items discriminated significantly between neurotic outpatients and normals (one item with P<0.05 and the rest with P<0.001). When self-ratings were compared (as opposed to counting check list responses) all items discriminated significantly (Kellner, 1972b).

In a crossover drug trial of diazepam, 5 mg t.i.d., hydroxyzine pamoate, 50 mg t.i.d. and placebo with anxious neurotics (N17), the check list (PP) discriminated significantly between the effects of treatments (P<0.005), whereas the check list form of the MAS and the Neuroticism Scale of the MPI failed to discriminate at a significant level (Kellner et al., 1968).

Correlations with other self-rating scales were similar to those of the total SRT scores which are listed below.

VALIDITY OF THE SRT (WEEK FORM) AND CORRELATIONS WITH OTHER TESTS
The literature of drug trials with neurotic patients was surveyed and those rating and self-rating scales which appeared to be effective in detecting differences between an active drug and placebo were selected (Kellner, 1971). These tests were compared with the SRT (TC or PP) in four drug trials. The results obtained with the SRT were similar to those obtained by psychiatrists' ratings using the Hamilton Anxiety Rating Scale.

<table>
<thead>
<tr>
<th>Subscales</th>
<th>Check list x SD</th>
<th>Total SRT score x SD</th>
<th>Anxiety x SD</th>
<th>Depression x SD</th>
<th>Inadequacy x SD</th>
<th>Somatic x SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinic 1 Houston (N50)</td>
<td>24.87 8.32</td>
<td>51.08 25.14</td>
<td>11.68 6.63</td>
<td>11.62 5.81</td>
<td>13.62 6.62</td>
<td>6.88 5.49</td>
</tr>
<tr>
<td>Clinic 2 Albuquerque (N50)</td>
<td>24.42 7.88</td>
<td>49.82 27.21</td>
<td>11.30 6.10</td>
<td>11.6 6.6</td>
<td>13.44 8.54</td>
<td>5.7 5.0</td>
</tr>
<tr>
<td>Normals 1 (N50)</td>
<td>7.02 5.69</td>
<td>8.98 8.33</td>
<td>1.5 1.9</td>
<td>1.5 2.1</td>
<td>2.26 2.82</td>
<td>1.6 1.8</td>
</tr>
<tr>
<td>Normals 2 (N50)</td>
<td>7.76 6.86</td>
<td>11.88 13.64</td>
<td>2.4 2.9</td>
<td>2.5 2.9</td>
<td>3.18 4.19</td>
<td>1.6 2.3</td>
</tr>
</tbody>
</table>

*The results were obtained with the original pencil and paper version; scores obtained with the short version tend to be lower.
TABLE 2
SENSITIVITY AND SPECIFICITY* AND MISCLASSIFICATION RATES\(^\d\) OF SRT SCORES (WEEK FORM, PENCIL AND PAPER) IN TWO STUDIES†

<table>
<thead>
<tr>
<th>Symptom Rating Test scores</th>
<th>Subscales</th>
<th>Taylor MAS</th>
<th>N Scale EPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Check list</td>
<td>Total SRT</td>
<td>Anxiety</td>
</tr>
<tr>
<td>Study 1 Predominantly new neurotic outpatients (N50) and normals (N50)</td>
<td>88</td>
<td>84</td>
<td>86</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>90</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>Specificity</td>
<td>11</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Misclassified‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2 (Neurotic outpatients) (N50) from another clinic and another group of normals (N50)</td>
<td>90</td>
<td>84</td>
<td>82</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>82</td>
<td>84</td>
<td>90</td>
</tr>
<tr>
<td>Specificity</td>
<td>14</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

See text.
*Results obtained with the Taylor Manifest Anxiety Scale and the Neuroticism Scale of the Eysenck Personality Inventory are listed for the purpose of comparison.
‡Patients who had low scores (below the cut-off point) or normals which had high scores (above the cut-off point).

(Hamilton, 1959), rating of Target Symptoms and global ratings (Kellner et al., 1968; Kellner and Sheffield, 1968; Kellner and Claghorn, 1970; Kellner, Gervais, and Pathak, 1972). The SRT was about as sensitive as psychiatrists' ratings in detecting drug/placebo differences (or differences between two drugs).

In seven studies the SRT (TC and PP) was administered to normal subjects and neurotic patients and all the scores discriminated significantly between the two populations (P<0.001). The SRT discriminated more accurately than the Neuroticism Scale of the MPI (or that of the EPI) and about as accurately as the MAS (Kellner, 1967a; Kellner and Sheffield, 1967; Morton, 1968; Kellner, 1972b).

Validation of the scoring method (TC) was carried out by comparing the scores of neurotic outpatients (N16) and normals (N24) who had the same number of symptoms. Neurotics had significantly higher total scores than normals even when the number of symptoms was the same (P<0.001) (Kellner and Sheffield, 1967).

The means and standard deviations of the SRT and subscales are listed in Table 1.

The SRT's sensitivity (how many cases the test fails to detect) and its specificity (how many normals it misclassifies as cases) (Reid, 1960) were compared with those of the MAS and the N Scale of the EPI. The three tests were administered to 50 neurotic outpatients and 50 normals, and this study was repeated in another outpatient clinic (N50) with another group of normals (N50). The specificity (the number of normals correctly identified by the test scores divided by the total number of normals) and the sensitivity (the number of patients correctly identified by the test scores divided by the total number of patients) are listed in Table 2. The largest number of patients and normals were misclassified by the test scores of the somatic subscale of the SRT, followed by the N Scale of the EPI. The MAS scores were similar to those of the total SRT scores and those of the other three SRT subscales, with the total SRT score and two of the SRT subscales showing a slightly smaller proportion of misclassified subjects. The rankings of specificity and sensitivity were similar to the misclassification rates; there were a few differences between rank orders of specificity and sensitivity of the tests in the two studies (Table 1).

Normal subjects who had been treated by their general practitioner with tranquillizers or
sedatives in the past five years (GP sub-group N11) had a significantly higher score (TC) than other normals (P<0.001) but lower than neurotic outpatients (P<0.001); the N Scale of the MPI did not discriminate significantly between the GP sub-group and the neurotic outpatients (Kellner and Sheffield, 1967).

There was a marked decrease in check list scores and total SRT scores (TC) in neurotic outpatients (N40) after one month's treatment (P<0.001). There was a further decrease in scores after two months, but the difference in scores after one month and two months was not significant. After two months there was no significant difference between the scores of neurotic outpatients and the scores of the GP subgroup of normals; the scores remained significantly higher than those of the total normal group (P<0.001).

Neurotic outpatients (N35) tested on their first attendance had significantly higher SRT (PP) scores (P<0.05) than neurotic inpatients and daypatients (N40), many of whom had apparently improved and were ready to be discharged (Kellner, 1967a).

In a longitudinal study with anxious schizophrenic outpatients (N20) there was a significant trend for SRT scores (PP) and psychiatrists' Hamilton Anxiety Scale scores to change in the same direction (P<0.01). The patients' SRT scores were significantly higher than those of normals (P<0.001).

SRT (PP) scores of new inpatients suffering from endogenous depression (N16) were compared with the scores of neurotic inpatients (N26) and those of normals (N51). Depressed patients had significantly higher SRT scores than neurotic patients (P<0.01). After three weeks' treatment with antidepressants their score had significantly decreased (P<0.001) but remained significantly higher than that of normals (P<0.01) (Kellner, 1972b).

SRT (PP) scores and subscales scores of alcoholic inpatients before treatment (N52) and after six weeks' treatment (N31) with group therapy and medication (largely minor tranquillizers) were compared with the scores of new neurotic outpatients (N72) and those of normals (N51). All SRT scores of alcoholic patients before treatment were significantly higher than those of normals (P<0.001); the total SRT score and the scores of the anxiety, depression, and inadequacy subscales were significantly lower than those of neurotic outpatients (anxiety, depression, and inadequacy with P<0.001, total SRT score with P<0.05). After treatment all scores of alcoholic patients were significantly lower than before treatment (P<0.001) and did not differ significantly from those of normals (Schenkel and Kellner, 1971).

The findings in a drug trial with anxious neurotics suggest that the Neuroticism Scale of the MPI and the MAS and the SRT (TC) differ in their stability (Kellner et al., 1968). The Neuroticism Scale tends to measure the personality trait of neuroticism (or emotionality) (Bendig, 1960) and the scores are more stable than those of the other two scales. The SRT tends to measure recently experienced distress, a temporary and changing state. The MAS appears to be intermediate between the two.

### TABLE 3

| Spearman Rank Order Correlation Coefficient (Rho) of SRT Scores (Week Form, Pencil and Paper), Taylor MAS Scores and Neuroticism Scores of the Eysenck Personality Inventory in Normal Subjects (N113) and Neurotic Outpatients (Predominantly New Patients, N126)* |
|---------------------------------|--------------|--------------|---------------|---------------|---------------|
|                                | Check list   | Total SRT    | Subscales     |               |               |
|                                |              | score        | Depression    | Anxiety       | Inadequacy    |
| Taylor MAS                     |              |              |               |               |               |
| Normals                        | 0.74         | 0.75         | 0.47          | 0.50          | 0.58          | 0.47          |
| Patients                       | 0.73         | 0.72         | 0.61          | 0.65          | 0.61          | 0.42          |
| Neuroticism Scale of EPI       |              |              |               |               |               |
| Normals                        | 0.70         | 0.70         | 0.47          | 0.60          | 0.70          | 0.41          |
| Patients                       | 0.73         | 0.71         | 0.61          | 0.60          | 0.56          | 0.51          |

*Correlations with other tests are listed in the text.
In a drug trial with neurotic patients suffering from anxiety reactions two psychiatrists rated the patients on four occasions using the Hamilton Anxiety Rating Scale. The Kendall Tau Correlation Coefficient (Siegel, 1956) between the SRT and the average ratings by two psychiatrists ranged from +0.58 to +0.83 (Kellner, 1967a).

In seven studies with neurotic patients and normals the correlation between the SRT scores (TC and PP) and the N Scale of the MPI (or EPI) ranged from +0.26 to +0.74, median +0.56. The correlations were significant except for one (+0.26, new neurotic outpatients, N40, TC) (Kellner, 1967a; Kellner and Sheffield, 1967; Morton, 1968; Kellner, 1972b).

The correlation between the SRT and the Zung Depression Self-rating Scale (Zung, 1965) for 25-13% of the variance. Factor IV—the MPI ‘introversion-extroversion’ factor which loaded around 0.80 and accounted for 11-12% of the total variance. The product moment correlation of factors I and III was +0.68 and were equally negatively related to extroversion (approximately −0.40). Factor II was independent (Bedford and Edington, 1968).

Positional set apparently played only a small part in the self-rating of neurotic symptoms in this study. The correlation for the average score per symptom obtained by randomized test cards and by the pencil and paper test of SRT was +0.90 and one week later was +0.95.

### OTHER FORMS

Several of the studies carried out with the Week form were repeated with the Day form and the results obtained with the two forms were similar (Kellner, 1967a; Morton, 1968).

Initially, there were a few differences between the cues of the self-rating scales of the Week and Day form. These differences were found to be unimportant and the Day form used in recent studies has been identical with the Week form except for the initial instruction (Appendix 2).

There is evidence to suggest that the average score of several self-ratings is more valid than a single self-rating (or psychiatrist’s rating) at the end of the treatment period. The average of several self-ratings is apparently less distorted by errors in recall (Kellner, 1971). In one drug trial the average Day scores significantly discriminated between the treatments, whereas the Week form and other ratings failed to do so (Kellner and Sheffield, 1968).

The Day form appears to be suitable when several measurements during a short treatment block are desired—for example, when changes from one day to the next are studied, such as in the assessment of early responses to treatment. The Day form also appears to be suitable for some forms of the ‘self-controlled trial’ (Hogben and Sim, 1953) or the ‘intensive design’ (Chassan, 1961) or the ‘multiple crossover design’—that is, the measurement of responses to repeated and systematic changes in treatment. In a single-patient multiple crossover drug study, in which the treatment periods lasted only a few days, the Day form (PP) discriminated effectively.
between the responses to the treatments (Kellner and Sheffield, 1968).

In a longitudinal study of 18 neurotic outpatients correlations between the SRT Day form (TC) and the total score of the Personal Questionnaire (Shapiro, 1961) were computed for each patient separately. The median correlation was +0.86. It appeared that the differences between the two tests—that is, the different techniques of self-rating, the individual scaling of cues and the number of cues available for the self-rating of each symptom—did not appreciably influence the changes in total scores. The SRT was less time-consuming and gave similar results to the PQ when changes in the total scores of neurotic patients were being measured (Sheffield and Kellner, 1970).

In several studies with neurotic patients the correlation between the SRT (PP and TC) check list, the SRT total score (Week and Day form), and the average SRT score per symptom, the Neuroticism scale of the MPI (and that of the EPI) and the MAS, were positive, and most correlations were significant. Thus, subjects with high N scores and high MAS scores tended to report having experienced more symptoms during the past week, and on the day of rating, and tended to rate these symptoms as more severe and more frequently occurring than patients with low N scores and low MAS scores.

Throughout the studies with the Week form and with the Day form, the Extroversion Scale of the MPI and EPI tended to be negatively correlated with the N Scale, the SRT and the MAS. In a few studies, the negative correlation between the SRT and the E Scale and between the MAS and the E Scale reached a significant level (Kellner, 1967a).

In several studies the ‘Hour’ Form was used—for example, in a study of diurnal variations in depressed patients (Kellner, 1967a). In other studies forms were used which focused on the more distant past such as ‘How have you felt during the past few years?’ Generally, the more recent the time focus—for example, ‘past hour’, ‘now’—the greater tendency for changes in scores from one test to the next, apparently reflecting fluctuations in the severity of distress in psychiatric patients. Forms focused on the more distant past (‘How have you felt during the past year’, ‘during the past few years?’) are more stable, tend to show a greater test–retest reliability, and are less suitable for the measurement of changes after a short time-interval such as one or two weeks.

| TABLE 5 |
| ITEMS OF SUBSCALES |

<table>
<thead>
<tr>
<th>Item no.</th>
<th>Anxiety</th>
<th>Item no.</th>
<th>Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Nervous</td>
<td>2</td>
<td>Tired, lack of energy</td>
</tr>
<tr>
<td>5</td>
<td>Scared, frightened</td>
<td>6</td>
<td>Poor appetite</td>
</tr>
<tr>
<td>9</td>
<td>Restless, jumpy</td>
<td>8</td>
<td>No hope</td>
</tr>
<tr>
<td>16</td>
<td>Trembling, shaking</td>
<td>12</td>
<td>Feeling guilty</td>
</tr>
<tr>
<td>19</td>
<td>Tense, wound up</td>
<td>18</td>
<td>Unworthy, failure</td>
</tr>
<tr>
<td>23</td>
<td>Thoughts you could not push out of your mind</td>
<td>24</td>
<td>Lost interest</td>
</tr>
<tr>
<td>26</td>
<td>Attacks of panic</td>
<td>25</td>
<td>Unhappy, depressed</td>
</tr>
<tr>
<td>29</td>
<td>Difficulty in falling asleep, restless sleep</td>
<td>30</td>
<td>Awakening early</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item no.</th>
<th>Somatic</th>
<th>Item no.</th>
<th>Inadequacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dizzy, faint</td>
<td>10</td>
<td>Poor memory</td>
</tr>
<tr>
<td>4</td>
<td>Feeling of pressure or tightness in head or body</td>
<td>13</td>
<td>Worrying</td>
</tr>
<tr>
<td>7</td>
<td>Heart beats quickly or strongly</td>
<td>15</td>
<td>People look down on you</td>
</tr>
<tr>
<td>11</td>
<td>Chest pains</td>
<td>17</td>
<td>Difficulty in thinking clearly or making up your mind</td>
</tr>
<tr>
<td>14</td>
<td>Muscle pains, aches</td>
<td>20</td>
<td>Inferior to others</td>
</tr>
<tr>
<td>21</td>
<td>Feeling numb or tingling</td>
<td>22</td>
<td>Irritable</td>
</tr>
<tr>
<td>27</td>
<td>Parts of body feel weak</td>
<td>28</td>
<td>Could not concentrate</td>
</tr>
</tbody>
</table>
TEST VALIDITY AND RESPONSE SET

The measurement of reliability and validity of a test designed to measure changes is somewhat different from that designed to measure a stable personality trait. The principles of the validation of a distress scale have been discussed elsewhere (Kellner, 1971a).

It was postulated that a valid test of changes in distress should (1) distinguish distressed patients from normals; (2) measure changes of distress comparable with those obtained by psychiatrists’ ratings; (3) discriminate between the responses of a drug which is known to reduce distress and the responses to placebo—that is, the chemical criterion (Clyde, 1960).

No attempt is made in the Symptom Rating Test to disguise the purpose of the ratings, and no suppressor or lie score is included. The SRT is likely to be influenced by the patient’s conscious or unconscious bias.

Response sets make a test less sensitive to changes, since a very strong response set would prevent changes altogether. It can be assumed that a test which is sensitive, and discriminates effectively between responses to different treatments.

Most response sets and biases might play an important part when differences between individuals are being determined. However, when changes are being measured, the biases and response sets of an individual are likely to remain the same unless it is assumed that the change of treatment or treatment itself could also change a person’s response sets. Two response biases have been postulated which might distort changes in self-ratings in experiments in therapies: the crossover drug trials the ‘residual effect’ (Raymond, Lucas, Beesley, O'Connell, and Roberts, 1957), but they appear to form only a small part of the total variability (Kellner, 1971).

COMPARISON OF PENCIL AND PAPER VERSIONS, TEST CARDS, AND OBSERVER RATING SCALES

There were no significant differences between the pencil- and paper-based tests and those obtained by observer rating scales in any of the studies, but there is some evidence to suggest that these methods may differ in their effectiveness.

The pencil and paper version and test cards were administered to two subgroups (each N15) of anxious neurotic patients who participated in two crossover studies with new investigational drugs. None of the rating and self-rating tests discriminated between the treatments at a significant level but in both studies there was a trend for the test card version to yield greater differences between the responses to drug and placebo. The reason for this difference may be that in the test card version the patient rated himself, whereas in the pencil and paper version he rated all symptoms contained in the checklist; the other reason may be that during the administration of the test cards the patients were more closely supervised and may have completed the test more conscientiously (Kellner, 1972b).

In drug trials with neurotic patients differences between the effectiveness of self-ratings and observer ratings have been small (Kellner, 1971, 1972a), but this finding apparently applies only to studies with groups of a certain size. If the groups who are illiterate, have poor eyesight, or for other reasons fail to complete their self-rating forms adequately, can confound differences between treatment effects. In these patients the SRT should be used as an observer rating scale (Appendix 2).

The Week form of the Short Version (PP) (Appendix 1) was compared with the original in two studies with neurotic outpatients and normals and in one parallel drug trial of benzoctamine and placebo (N70). The differences between classification rates of the two versions were small. Similarly, differences between the changes in scores during the drug trial were small and there appeared no advantage in using the more time-consuming original version (Kellner, 1972b).

The results of the studies in which various versions of the SRT were compared are inconclusive. Their effectiveness in discriminating between distressed patients and normals appeared to be similar; further studies are likely to reveal whether there is a substantial difference.
between the responses to psychotropic drugs and placebo.

**PRACTICE WEEKS**

There is a large reduction in distress scores soon after attendance at a clinic even before a placebo is given (Gliedman, Nash, Imber, Stone, and Frank, 1958; Frank, Nash, Stone, and Imber, 1963; Kellner and Sheffield, 1971). This initial improvement in apparently overshadow specific effects of psychotherapy (Kellner, 1967b), and it can confound drug/placebo differences (or differences between two drugs) if the differences are small and if the customary number of patients are included in the trial. The findings in one drug trial and two studies with the SRT suggest that this confounding effect can be reduced by using ‘practice weeks’—that is, a time-interval consisting of testing and treatment between the first attendance at a clinic (or admission to hospital) and the beginning of the experiment (Kellner, 1967a; Kellner and Sheffield, 1968; 1971).

We are very grateful to all our colleagues who have helped us throughout this project, mainly to Dr. Dennis Bromley for his advice. We wish to thank the Mental Health Research Fund and the Liverpool Regional Hospital Board for financial support which made this work possible.

**REFERENCES**


Self-rating scale


