
The Bay Area Functional Performance Evaluation: Development and Standardization

Dale Houston, Susan Lang Williams,
Judith Bloomer, William C. Mann

Key Words: activities of daily living •
mental health • task performance and
analysis • tests, by title, psychosocial, BaFPE

The Bay Area Functional Performance Evaluation (BaFPE) was developed in 1977–1978 to meet the need for a reliable and valid instrument for assessing the general functional performance of patients treated in psychiatric occupational therapy. It consists of two subtests, the Task-Oriented Assessment and the Social Interaction Scale. These subtests evaluate two aspects of general functional performance—task-oriented and social behavior—that are important in assessing clients with emotional, cognitive, or behavioral deficits. This article traces the instrument's standardization over a 10-year period of development and includes a discussion of its theoretical premises, its content, and the revisions to date. Research on the reliability and validity of the BaFPE is summarized.

Dale Houston, MA, OTR, FAOTA, at the time of this study, was Chief of Rehabilitation Services, Langley Porter Psychiatric Institute, University of California at San Francisco. (Mailing address: 804 Marylyn Circle, Petaluma, California 94952.)

Susan Lang Williams, MBA, OTR, ATR, is Work Projects Consultant, Community Mental Health Services, San Francisco, California.

Judith Bloomer, PhD, OTR/L, is employed at the Center for Forensic Psychiatry, Hamilton, Ohio.

William C. Mann, PhD, OTR, FAOTA, is Director of Graduate Studies and Associate Professor, State University of New York at Buffalo.

This article was accepted for publication June 14, 1988.

The Bay Area Functional Performance Evaluation (BaFPE) was developed in 1977–1978 by Bloomer and Williams (1979) to meet the need for a reliable and valid instrument for assessing general functional performance of patients treated in psychiatric occupational therapy. This paper presents information about the development of both the original and the revised BaFPE and the progress of its standardization.

Theoretical Principles

The development of the BaFPE was stimulated by the desire to evaluate psychiatric patients' functioning in general day-to-day activities. The authors focused on the concrete verbal and nonverbal behavior that can be seen to have a purpose in relating to objects and people in the environment. This orientation is based on the principle that skills are learned, acquired, and then actively produced in response to demands of the environment. This principle has been thoroughly formulated as guiding theory in the development of the occupational therapy profession, in particular as the acquisitional frame of reference (Mosey, 1970) and the occupational behavior frame of reference (Moorhead, 1969; Reilly, 1966). The principle that activity is adaptive to the environment is also a major tenet in the history of behavioral psychology.

Behaviorism grew out of the functionalist movement of the early 1900s (Herrnstein & Boring, 1965; Watson, 1913). Functional behavior, from a strictly behavioral standpoint, is concerned with the manifestation of psychological processes in observable behavior. The functionalist point of view was close to common sense, having to do with success in living and the adaptation of the person to his or her environment. Functional psychology looks for functional relations—"for the dependence of this on that" (Boring, 1957, p. 560).

These functional relations were investigated by the developers of the BaFPE. In terms of adaptive action, the authors asked what processes are required for functional performance in day-to-day living. What is necessary to complete a task, to perform an activity, to engage in social interaction? On what functional parameters are the activities dependent? These functional behaviors, a basic requirement for fundamental activity, are viewed as precursors to occupational behavior.

In recent years, the Model of Human Occupation has been given much attention in occupational therapy literature (Kielhofner, 1985; Kielhofner & Burke, 1980; Kielhofner, Burke, & Igi, 1980). This model has been seen by Reed (1984) and Robertson (1985) as well as by the BaFPE authors as the synthesis of four existing frames of reference for occupational therapy: adaptive performance, biodevelopment, facilitation of

growth and development, and, particularly, Reilly's (1966) occupational behavior. In the Model of Human Occupation, occupational behavior is described as an individual's ability to competently perform everyday tasks and behaviors in work, play, and self-care, and to attain satisfaction from these activities. Underlying these abilities are interpersonal communication/interaction skills, process skills for dealing with events (e.g., problem solving, planning), and perceptual-motor skills involved in responding to and acting on the environment (Barris, Kielhofner, & Watts, 1983). The occupational therapy process, which is part of this model, emphasizes assessment of these component skills as a basis for evaluation and treatment planning. The BaFPE focuses on the functional parameters that underlie these component skills.

In summary, *functional behavior* is defined as the employment of useful activity to achieve an active mode of adaptation to everyday living situations. This useful activity is possible because of skills acquired throughout life and promotes successful interaction with people and objects in the environment. Thus, the BaFPE is consistent with a functionalist perspective and with the acquisitional and occupational behavior frames of reference in that it focuses on present functioning, evaluates functional parameters reflected by acquired skills, and assumes that successful interaction with the environment includes the productive and active use of these skills. In the original BaFPE, these skills were measured through a sampling of task and social behaviors.

The Original BaFPE

The BaFPE was developed over a 1½-year period (1977–1978) at Langley Porter Psychiatric Institute in San Francisco. The authors of the test, Bloomer and Williams (1979), used the theoretical principles discussed above to develop a valid and reliable instrument for assessing the general functional performance of patients treated in psychiatric occupational therapy. They believed that successful adaptation to the environment would involve two general abilities: (a) the ability to engage in goal-directed and task-oriented interaction with objects in the environment and (b) the ability to interact in a socially appropriate way with people in the environment. The BaFPE assesses performance in the first area, goal-directed activity, through the Task-Oriented Assessment, and performance in the second area, socially appropriate behavior, through the Social Interaction Scale.

The Task-Oriented Assessment (TOA)

In the original TOA, five tasks were used to rate 10 different functional behaviors. (The process of task

selection began with the consideration of 10 tasks. After evaluation and preliminary testing, it was found that five tasks were sufficient to evaluate the chosen areas of functional behavior.) The content of the five tasks eventually chosen did not reflect essential life skills but provided a range of structure that enabled clinicians to evaluate functional parameters underlying the task-oriented behavior of their clients. Briefly, the five tasks were as follows:

1. Sorting Shells—Sorting 10 categories of shells by size, shape, and color.
2. Bank Deposit Slip—Following written instructions for filling out a bank deposit slip that involved several mathematical calculations.
3. House Floor Plan—Drawing a floor plan for a house with specific instructions about what it should include.
4. Block Design—Duplicating a block design by memory, or with the assistance of a cue card if needed.
5. Draw a Person—Drawing a picture of a person doing something.

The following 10 functional parameters were rated during the five tasks (see definitions in Bloomer & Williams, 1979, pp. 27–30):

1. Ability to paraphrase instructions
2. Ability to make productive decisions
3. Level of motivation
4. Ability to organize time and materials
5. Degree of mastery and self-esteem
6. Level of frustration tolerance
7. Attention span
8. Ability to abstract
9. Evidence of thought or mood disorder (verbal or behavioral)
10. Ability to follow instructions, leading to correct task completion

The clinicians observed and rated each task along a functional continuum encompassing the following broad delineations of behavior: *markedly dysfunctional or inappropriate* (1), *somewhat dysfunctional or inappropriate* (2), *usually functional or appropriate* (3), and *almost always functional or appropriate* (4).

In addition, space was provided on the rating sheet for clinicians to note the presence of various behaviors that denoted intact perceptual-motor functioning.

The Social Interaction Scale (SIS)

The original SIS was a rating scale that assessed the behavior of an individual in interactions with others in a social setting. Seven different components of social

interaction were evaluated. In each of these the authors sought to conform to the social values of American society as a whole. For example, taking direction from authority figures is considered a reality to be dealt with, and response to such direction should be appropriate, nonhostile, and functionally accommodating, relative to the situation at hand.

The following seven components of social interaction were assessed (see definitions in Bloomer & Williams, 1979, pp. 39-42):

1. Response to authority figures
2. Verbal communication
3. Psychomotor behavior
4. Independence/dependence
5. Socially appropriate behavior
6. Ability to work with peers
7. Participation in group and/or program activities

These components of the SIS were rated, as were those of the TOA, along a continuum outlining broad delineations of functioning. The 4-point rating scale used on the TOA was replaced by a 5-point rating scale to allow for rating a client who was too ill to interact with others in a group setting. Specific observation settings were not identified, but it was suggested that a client be observed in at least two different types of social situations: a more structured setting and a less structured setting. No specific time frame for making the observations was established.

The original BaFPE, called the research edition, was first published privately by Bloomer and Williams in 1979. This manual provided detailed information on the development of the BaFPE and on the theoretical principles on which it was based, as well as information on general research methodology. In addition,

it reviewed two case studies in which the BaFPE was used. Subsequently published commercially (Bloomer & Williams, 1979), it achieved wide distribution throughout the United States and by 1983 was beginning to be used in Canada. The authors presented workshops on the BaFPE in many states and at local, state, and national conferences. Between 1979 and 1983 a field study was conducted that provided much valuable feedback from clinicians about both the TOA and the SIS. In particular, the field study unearthed some problems that pointed to the need for a revision of the BaFPE: In the TOA, some of the tasks were perceived to be economically and geographically biased, and some of the instructions and rating guidelines were ambiguous. In the SIS, the failure to specify social settings in which the ratings were to be made caused difficulties. In 1983, a revision was begun.

The Revised BaFPE

The revision of the BaFPE was a long process that eventually took as long as the development of the original instrument. The revision involved analyzing large amounts of clinical feedback regarding content and scoring, assessing statistician feedback after data analysis, compiling the revision itself, and developing a new study to test the reliability of the revised instruments. Some of the content and structure of the original BaFPE was altered in the revision to reduce bias, improve the clarity of the rating guidelines, and simplify the scoring format. Table 1 summarizes the changes made in the revision.

The TOA

Two tasks, the Bank Deposit Slip and the House Floor Plan, were revised substantially, but all five tasks un-

Table 1
Summary of BaFPE Changes

	Original BaFPE	Revised BaFPE
TOA/SIS:	Scores combined to yield a composite BaFPE score.	TOA/SIS: Scored separately to yield a task score and a social interaction score.
TOA/SIS:	No reporting format provided.	TOA/SIS: Reporting format provided for each.
TOA:	Ratings for all tasks done on two behavioral guideline sheets.	TOA: Ratings done on separate sheets for each task.
TOA:	Bank Deposit Slip task thought to contain possible economic bias.	TOA: Money and Marketing task substituted for Bank Deposit Slip task.
TOA:	House Floor Plan task thought to contain possible cultural/geographic bias.	TOA: Home Drawing task has clarified instructions and revised list of rooms.
TOA:	Ten functional parameters.	TOA: Twelve functional parameters grouped into three component areas: Cognitive, Affective, and Performance.
TOA:	Perceptual-motor observations made on summary score sheet and added to the score.	TOA: More formalized section, Qualitative Signs and Referral Indicators, focuses on gross screening for organicity and is not formally scored.
SIS:	No specified social settings on which ratings were to be based.	SIS: Delineates five specific rating settings on which scores are to be based.
SIS:	Provided no means by which clients could assess their own social skills.	SIS: Provides optional self-report of social interaction to be completed by the client.

Note. BaFPE = Bay Area Functional Performance Evaluation. TOA = Task-Oriented Assessment. SIS = Social Interaction Scale.

MONEY AND MARKETING TASK: RATING SHEETS	
(Time limit: 8 minutes)	
Client's name _____	Date _____
Examiner _____	
COGNITIVE COMPONENTS	
Memory for Written/Verbal Instructions	
Definitions	General Idea: Buy some items and figure out how much money is left.
	Important details: 1. Will receive a list of items (groceries) to buy. 2. Will receive money (cash and check). 3. If needed, to buy items, will endorse and cash check. 4. Figure out how much money is left after buying items.
{ }	1. Not able to restate general idea even with one repetition of instructions.
{ }	2. Restates general idea with one repetition of instructions. May or may not include all four details.
{ }	3. Restates general idea without repetition of instructions. Does not include all four details.
{ }	4. Restates general idea without repetition of instructions and includes all four details.

Figure 1. Example portion of Task-Oriented Assessment rating sheet: Money and Marketing Task. From *Bay Area Functional Performance Evaluation* (2nd ed.) (p. 34) by S. L. Williams and J. Bloomer, 1978 and 1987, Palo Alto, CA: Consulting Psychologists Press, Inc. Copyright 1978 and 1987 by Judith Bloomer and Susan Williams. Reprinted by permission. Further reproduction is prohibited without the publisher's consent.

derwent at least minor revision to clarify rating instructions or behavior-rating guidelines. The Money and Marketing task, which replaces the Bank Deposit Slip task, retains the processes of the former task, but incorporates them in a more common activity. The instructions are written as before, and the task requires the subject to copy information, calculate the cost of items on a shopping list, determine if there is sufficient money to complete the purchase or if the supplemental check must be cashed, and calculate the amount of change left over. The Home Drawing task, which modifies the House Floor Plan, has more specific instructions, a list of more typical rooms, and a different sample of floor plans.

One of the 10 functional parameters was eliminated as redundant, and Parameter 10, the Completion parameter, was expanded into 3 parameters to separate the quantitative, qualitative, and time efficiency aspects of task completion. The resulting 12 parameters were then grouped as follows:

- Cognitive Component**
 - Memory for Written/Verbal Instructions
 - Organization of Time and Materials
 - Attention Span
 - Evidence of Thought Disorder
 - Ability to Abstract
- Performance Component**
 - Task Completion
 - Errors
 - Efficiency
- Affective Component**
 - Motivation or Compliance
 - Frustration Tolerance
 - Self-Confidence
 - General Affective and Behavioral Impression

Figure 1 presents a portion of the rating sheet for the Money and Marketing task, the portion used to assess the first cognitive parameter, Memory for Written/Verbal Instructions.

The Qualitative Signs and Referral Indicators section of the revised TOA includes items that allow the rater to check for general signs of an organic disorder that are observable during all five tasks. These signs cover five areas: (a) expressive language, (b) comprehension of written and auditory language, (c) hemispatial neglect, (d) memory, and (e) abstraction. In addition, this section includes items that are specific to each task. Figure 2 shows an example of the task-specific observations for the Money and Marketing task, which reflect the written instructions and arithmetic operations unique to this task. The Qualitative Signs section is not included in the scoring of the TOA, however. The number of signs noted is tallied and if the total exceeds a specified number, further testing for organic disorders may be indicated.

The SIS

The revised SIS uses the same seven parameters of social functioning and the same 5-point rating scale as did the original SIS. The major change is that the SIS scores are now based on a series of ratings made in five specified social situations, rather than on the original single rating done after observing the patient in whatever social situations were available. This change was intended to provide a more objective sampling of

QUALITATIVE SIGNS AND REFERRAL INDICATORS										
Check if symptom is present. Starred items indicate higher potential for organic involvement and need for further testing (see TOA Appendix C).										
General Observations (may duplicate areas already scored)										
Task-Specific Observations										
{ }	1. Not able to read instruction sheet: visual impairment due to not having appropriate glasses, blurred vision.									
{ }	*2. Not able to read instruction sheet: etiology unknown.									
{ }	3. Not able to read instruction sheet: not able to read, illiterate.									
{ }	4. Reading disorder: inability to read less familiar words/multisyllable words.									
{ }	*5. Reading disorder: reads with letter reversals.									
{ }	6. Reading disorder: inability to comprehend sentence meaningfully although words are OK (seen in "memory for instruction" question).									
{ }	*7. Calculation error: does not differentiate math processes (e.g., adds everything)									
{ }	*8. Calculation error: number reversals cause errors.									
{ }	*9. Calculation error: errors in place holding (e.g., writes \$9.15 as \$900.15).									
{ }	*10. Calculation error: uses segmental addition or subtraction, i.e., fails to carry number to next column.									
Example:	<table style="display: inline-table; vertical-align: middle;"> <tr> <td style="text-align: right;">23</td> <td style="padding: 0 20px;">instead of:</td> <td style="text-align: right;">23</td> </tr> <tr> <td style="text-align: right;">+ 69</td> <td></td> <td style="text-align: right;">+ 69</td> </tr> <tr> <td style="text-align: right;">812</td> <td></td> <td style="text-align: right;">92</td> </tr> </table>	23	instead of:	23	+ 69		+ 69	812		92
23	instead of:	23								
+ 69		+ 69								
812		92								

Figure 2. Example portion of task-specific observation: Money and Marketing Task. From *Bay Area Functional Performance Evaluation* (2nd ed.) (p. 35) by S. L. Williams and J. Bloomer, 1978 and 1987, Palo Alto, CA: Consulting Psychologists Press. Copyright 1978 and 1987 by Judith Bloomer and Susan Williams. Reprinted by permission. Further reproduction is prohibited without the publisher's consent. (Note. The *General Observations* portion of this form is not shown here.)

social behavior. The five social situations specified are the following:

1. One-to-One—usually rated during the examiner's initial interview.
2. Mealtime—general group eating situation.
3. Unstructured Group—such as informal lounge group or informal recreational activity.
4. Structured Group Activity—an organized task group or sports or recreation event involving teams, partners, etc.
5. Structured Verbal Group—an organized group emphasizing discussion of current events, therapy group, community meeting, etc.

Figure 3 shows four of the seven parameters. It will be seen that some parameters cannot be rated in all situa-

BaFPE SOCIAL INTERACTION SCALE
Susan Lang Williams and Judith S. Bloomer

Client's name _____ Rater _____ Date _____

SITUATION SCORES (Circle Number)					PARAMETERS AND BEHAVIORAL GUIDELINES					PARAMETER SCORES		
One-to-One	Mealtime	Unstructured Group	Structured Activity or Sports	Structured Verbal Group						Total of the circled scores =	divided by	
1	1	1	1	1	I. VERBAL COMMUNICATION (Note if client's primary language is not that spoken in the treatment setting _____.)					13		5
2	2	2	2	2	1. Not able to assess due to degree of dysfunction _____, or due to language barrier _____.					5		5
3	3	3	3	3	2. Avoids verbal interaction or verbal interaction is inappropriate.					5		5
4	4	4	4	4	3. Verbal interactions appropriate only when directly questioned.					5		5
5	5	5	5	5	4. Able to sustain a logical conversation with others some of the time.					5		5
5	5	5	5	5	5. Will initiate or sustain a logical conversation most of the time.					5		5
					II. PSYCHOMOTOR BEHAVIOR					13		5
1	1	1	1	1	1. Not able to assess due to degree of dysfunction. Note if on meds _____.					5		5
2	2	2	2	2	2. Is markedly immobilized _____, or is hyperactive (with agitated behavior, pressured speech).					5		5
3	3	3	3	3	3. Is somewhat immobilized or hypoactive _____, or is somewhat hyperactive in verbal or motoric behavior _____.					5		5
4	4	4	4	4	4. Usually maintains a level of activity appropriate to situation.					5		5
5	5	5	5	5	5. Almost always maintains a level of activity appropriate to situation.					5		5
					III. SOCIALLY APPROPRIATE BEHAVIOR					20		5
1	1	1	1	1	1. Not able to assess due to degree of dysfunction.					5		5
2	2	2	2	2	2. Markedly inappropriate (e.g., very intrusive, hostile, or blatantly psychotic behavior in this social setting).					5		5
3	3	3	3	3	3. Somewhat inappropriate: general behavior in this situation (tests limits of social acceptability).					5		5
4	4	4	4	4	4. Usually appropriate in this situation (inappropriate behavior confined to one topic or area).					5		5
5	5	5	5	5	5. Almost always appropriate behavior in this setting.					5		5
					IV. RESPONSE TO AUTHORITY FIGURES (therapist, nurse, doctor, etc.)					13		5
1	1	1	1	1	1. Not able to assess due to degree of dysfunction.					5		5
2	2	2	2	2	2. Markedly inappropriate or negative response to authority.					5		5
3	3	3	3	3	3. Somewhat inappropriate: avoids contact with authority figures, or interaction tests limits of relationship (defiant, challenging, etc.).					5		5
4	4	4	4	4	4. Usually appropriate, except around specific areas.					5		5
5	5	5	5	5	5. Almost always appropriate interaction with authority figures.					5		5
11	4	8	14	11	TOTAL SCORE SIDE ONE							

Figure 3. Portion of the Bay Area Functional Performance Evaluation (BaFPE) Social Interaction Scale. From *Bay Area Functional Performance Evaluation* (2nd ed.) (p. 64) by S. L. Williams and J. Bloomer, 1978 and 1987, Palo Alto, CA: Consulting Psychologists Press. Copyright 1978 and 1987 by Judith Bloomer and Susan Williams. Reprinted by permission. Further reproduction is prohibited without the publisher's consent. (Note. The ratings marked on the form are explained in the book.)

Table 2
Summary Interrater Reliability for Original BaFPE

Scale	Original Study r^{11}	Field Study Range of r^{11}
TOA total	.99	.86-.99
SIS total	.83	.69-.91
BaFPE total	.99	.82-.97

Note. BaFPE = Bay Area Functional Performance Evaluation. TOA = Task-Oriented Assessment. SIS = Social Interaction Scale.

¹¹ Pearson product-moment correlations.

tions (see Parameter 4, Response to Authority Figures). The revised SIS includes an optional Self-Report of Social Interaction to be completed by patients that provides information about their insight regarding their own social functioning. However, this self-report is not used in scoring.

Standardization of the BaFPE

The Original BaFPE: Reliability

When the BaFPE was originally developed and tested in 1978 on 62 patients and 20 nonpatients, it showed excellent interrater reliability (Bloomer & Williams, 1979). In 1981, a field study was conducted (a) to determine whether the BaFPE could be reliably administered by occupational therapists using just the instruction manual and (b) to obtain feedback from users. Teams of therapists from seven hospitals in different parts of the United States participated in this study, each testing groups of 7 to 20 patients or nonpatients. Four teams tested a total of 51 patients, and three teams tested a total of 50 nonpatients. Interrater reliability results were generally very good, although they were not as high as in the original study and varied somewhat from center to center (see Table 2).

The Revised BaFPE: Reliability

Both the TOA and the SIS underwent enough revision to warrant another test of reliability. This was designed to be accomplished in two phases: (a) a qualitative evaluation of the revised instruments and (b) an interrater reliability study.

Qualitative evaluation of the revised BaFPE.

This study was conducted to make certain that the instructions for administering, rating, and scoring the revised instruments were clear and unambiguous and that the new formats of the rating and scoring sheets were clear and easy to use. Eight psychiatric occupational therapists reviewed the new manual, forms, and accessory materials. They administered the revised BaFPE to two or three persons, including at least one patient, to provide feedback about problems that emerged in use. Some suggestions for minor changes were incorporated into the final version of the BaFPE used in the interrater reliability study itself.

The interrater reliability study: Methodology. This study essentially replicated the reliability studies done with the original BaFPE. The raters were four pairs of California registered occupational therapists who worked in four acute psychiatric hospitals, three in the San Francisco Bay Area and one in Orange County. They learned to administer the TOA and SIS by using the revised instruction manual.

Each research team was to evaluate 25 patients who were at least 16 years old and had a specified DSM-III diagnosis (American Psychiatric Association, 1980) of an acute psychotic disorder. The patients were included in the order they were admitted to the hospital.

For each patient, the TOA and SIS were both completed within 7 days following the patient's admission to the hospital and within 24 hours of each other. Each member of the research team administered and rated the TOA for about half the subjects. In addition, the TOA administrator was responsible for orienting each subject to the research, obtaining consent to participate in the study, and interviewing the subject to obtain demographic data. The second therapist observed and rated the subject's performance on the TOA. For the SIS, both researchers observed each subject simultaneously in the same five settings, but independently rated the subject's social interactions. The TOA administrator was responsible for seeing that the subject completed and returned the Self-Report of Social Interaction.

The interrater reliability study: Results. Ninety-one patients from the four hospitals were included in the study. Data were obtained for 18 to 25 patients per hospital. Demographic characteristics differed from group to group, particularly in sex, ethnicity, and employment status at admission. In addition, the distribution of diagnoses varied from group to group, particularly in the relative proportions of schizophrenia to other types of psychosis. Summary demographic data for the four groups can be found in Williams and Bloomer (1987). Data analyses were done separately for the four groups, and where an overall indicator was wanted, an unweighted average of the correlations or means was calculated.

Reliability of the TOA. The TOA underwent more structural revision than the SIS, and the hypothesis that the reliability of the TOA would be improved by the revision was borne out by results. The following aspects of reliability were studied: (a) interrater reliability by group, (b) effects of changes in the instrument on interrater reliability, and (c) internal reliability.

The results of the analysis of interrater correlations across groups, using Pearson product-moment correlations, were generally very strong, particularly for the Performance component items, which are the

most concrete in nature (see Table 3). The lower range of correlations for the Affective component was expected because items in this component require more subjective judgment on the part of the rater than do most items in the other two components.

The task scores with the highest and most consistent correlations were Kinetic Person Drawing and Block Design. The task showing the most variability of correlation was Sorting Shells. The two changed tasks, Money and Marketing and Home Drawing, each showed three correlations over .90 and one lower correlation. In summary, over 80% of the correlations for the four groups equaled or exceeded .80, but there were evident differences in correlations across groups as well as within groups.

The second examination of reliability was a comparison of the correlations for items changed in the revised TOA with correlations for similar items in the original TOA. A summary of the results is shown in Table 4. For the new items—the three component scores, and the parameters Errors, Efficiency, and General Affective Impression—the level and range of correlations were very good, particularly for the Cognitive and Performance component scores. For the revised parameters, the range of correlations was found to be generally higher and tighter than before: Memory for (written/verbal) Instructions (.86-.94 compared with .63-.98), Task Completion (.87-.98 compared with .45-.99), and Self-Confidence (.74-.97 compared with .50-.97). The correlations for the parameter Ability to Abstract were lower in the revision than in the original (.68-.92 compared with .71-1.00). This may have been due to an error on the rating sheets that was later corrected. The two tasks revised to reduce the chance of cultural bias—Money and Marketing and Home Drawing—showed little change from the field study, although there was a slight reduction in the strength of the correlations for Money and Marketing (.66-.97 compared with .77-1.00). Still, the TOA total correlation, which was already quite high in the original BaFPE, improved further with this revision (.93-.98 compared with .86-.97).

In summary, correlation coefficients for 10 of the original 16 scales on the TOA improved with the revised instrument, 3 were lower, and 3 remained about the same. The items added to the revised instrument showed high correlations.

The third aspect of reliability studied was the internal reliability of the TOA. This aspect was studied to determine the degree to which the subscales correlated with each other in measuring the general construct of functioning in the TOA setting. A correlation matrix for rater scores was computed using the Pearson product-moment correlation procedure. The intercorrelations within each component area ranged

Table 3
TOA Interrater Correlations by Hospital Group

	Group 1 (N = 18)			Group 2 (N = 25)		Group 3 (N = 23)		Group 4 (N = 25)	
	n ^a	r	p*	r	p*	r	p*	r	p*
Cognitive Parameters									
Memory for Instructions		.86		.94		.91		.92	
Organization	14	.92		.92		.93		.80	
Attention Span	17	.64	.01	.96		.96		.81	
Evidence of Disorder	16	.73		.96		.83		.61	
Ability to Abstract	13	.89		.92		.77		.68	
Cognitive component	12	.93		.98		.92		.87	
Performance Parameters									
Completion	15	.92		.97		.98		.87	
Errors	15	.66	.01	.91		.91		.97	
Efficiency	17	.94		.90		.90		.93	
Performance component	13	.94		.97		.97		.97	
Affective Parameters									
Motivation and Compliance		.81		.86		.69	.005	.63	
Frustration Tolerance		.49	.05	.92		.83		.80	
Self-Confidence		.74		.97		.82		.75	
General Affective Impression		.66	.005	.86		.89		.66	
Affective component		.72		.96		.87		.85	
Task totals									
Sorting Shells	17	.75		.94		.78		.86	
Money and Marketing	16	.66	.005	.97		.95		.92	
Home Drawing	12	.95		.98		.53	.01	.94	
Block Design	16	.96		.97		.90		.90	
Kinetic Person Drawing	16	.93		.97		.93		.94	
Total score		.93		.98		.97		.94	

Note. From *Bay Area Functional Performance Evaluation* (2nd ed.) (p. 10) by S. L. Williams and J. Bloomer, 1978 and 1987, Palo Alto, CA: Consulting Psychologists Press. Copyright 1978 and 1987 by Judith Bloomer and Susan Williams. Adapted by permission. Further reproduction is prohibited without the publisher's consent. TOA = Task-Oriented Assessment.

^a n shown where number of cases used is less than full sample.

* p values are two-tailed and are $p \leq .001$ unless otherwise indicated.

from .45 to .80 for the Cognitive component, from .67 to .83 for the Performance component, and from .57 to .84 for the Affective component. There was at least one very high correlation within each component. Organization of Time and Materials correlated with Evidence of Thought Disorder at .80, Errors with Efficiency at .83, and Frustration Tolerance with Self-Confidence at .84, but these appeared to reflect key relationships rather than redundant scales. For example, since half of the Efficiency score is based on the Errors score, a high correlation was to be expected.

The intercorrelations for the three component areas were .56 for Affective/Performance, .69 for Cognitive/Performance, and .79 for Cognitive/Affective. The high correlation between the Cognitive and Affective components was not expected. However, examination of the correlations between the parameters making up these two components revealed that the major contribution to this high correlation appeared to arise from the interaction of Attention Span with Frustration Tolerance ($r = .80$), Attention Span with Self-Confidence ($r = .75$), and Organization of Time and Materials with Frustration Tolerance ($r =$

.74). It seems reasonable that ability to tolerate frustration would significantly influence a person's ability to attend to and organize performance.

The five tasks showed modest intercorrelations, indicating that they each contribute independent information to the functional evaluation. The Money and Marketing task showed the highest intercorrelations (.53 to .63) and the Kinetic Person Drawing the lowest (.29 to .53).

Reliability of the SIS. As described above, the main change in the revised SIS was scoring based on sample ratings in five specified social situations instead of on the former single global rating. All four teams rated all their patients in the one-to-one situation, but only Groups 3 and 4 completed enough paired ratings in the four group situations for the results to be included in the data analysis. The following aspects of reliability were evaluated: (a) interrater reliability, (b) effects of multiple observation situations on interrater reliability, (c) internal reliability, and (d) correlation between Self-Report of Social Interaction and SIS ratings.

Table 5 presents the interrater correlations for

Table 4
Comparison of Interrater Correlations for the Original^a and Revised TOA

Original Items	Mean <i>r</i> ^c	Range of <i>r</i>	Revised/New Items	Mean <i>r</i> ^c	Range of <i>r</i>
Paraphrase	.84	.63-.98	Memory for Instructions	.91	.86-.94
Decision Making ^b	.80	.39-.98	—	—	—
Organization	.89	.66-.98	Organization	.89	.80-.93
Attention Span	.95	.86-1.00	Attention Span	.84	.64-.96
Evidence of Disorder ^b	.76	.51-.85	Evidence of Disorder	.78	.61-.96
Ability to Abstract	.95	.71-1.00	Ability to Abstract	.82	.68-.92
—	—	—	Cognitive component	.93	.88-.93
Task Completion ^b	.91	.45-.99	Task Completion	.94	.87-.98
—	—	—	Errors	.86	.66-1.00
—	—	—	Efficiency (time and errors)	.92	.90-.97
—	—	—	Performance component	.96	.94-.97
Motivation	.79	.44-.93	Motivation/Compliance	.72	.63-.86
Frustration Tolerance	.75	.50-.94	Frustration Tolerance	.75	.49-.92
Sense of Mastery ^b	.67	.50-.97	Self-Confidence	.80	.74-.97
—	—	—	General Affective Impression ^b	.74	.66-.89
—	—	—	Affective component	.85	.72-.96
Sorting Shells	.86	.43-.99	Sorting Shells	.81	.75-.94
Check Deposit Slip ^b	.90	.77-1.00	Money and Marketing	.87	.66-.97
House Floor Plan ^b	.85	.57-.99	Home Drawing	.84	.53-.98
Block Design	.92	.84-.97	Block Design	.93	.90-.97
Draw a Person	.86	.76-.98	Kinetic Person Drawing	.94	.93-.97
TOA total	.94	.86-.97	TOA total	.96	.93-.98

Note. From *Bay Area Functional Performance Evaluation* (2nd ed.) (p. 12) by S. L. Williams and J. Bloomer, 1978 and 1987, Palo Alto, CA: Consulting Psychologists Press. Copyright 1978 and 1987 by Judith Bloomer and Susan Williams. Adapted by permission. Further reproduction is prohibited without the publisher's consent. TOA = Task-Oriented Assessment.

^a Field study with seven rater teams. ^b Items with substantive changes. ^c Unweighted means of group correlations.

the original and revised SIS parameters and for the observation situations of the revised SIS. The correlations for the SIS were generally lower than those for

the TOA in this study, but this outcome was similar to results in the previous studies with the original BaFPE, and was expected because the SIS parameters

Table 5
Social Interaction Scale: Comparison of Interrater Correlations for Original and Revised Instrument

	Original SIS ^a (N = 94)					Revised SIS				
	<i>n</i> ^a	Range of <i>r</i> ^b	<i>p</i> [*]	<i>r</i> ^b	<i>p</i> [*]	Group 3 (N = 20)			Group 4 (N = 25)	
						<i>n</i> ^a	<i>r</i>	<i>p</i> [*]	<i>r</i>	<i>p</i> [*]
Parameters										
Verbal Communication		.45-1.00		.69		.94		.83		
Psychomotor Behavior		.50-.82		.68		.77		.70		
Socially Appropriate Behavior		.45-1.00		.54		.85		.73		
Response to Authority Figures	60	.35-.79		.65		.96		.76		
Independence/Dependence		.52-1.00		.71		.79		.75		
Ability to Work With Others		.44-.85		.64		.66	.005	.77		
Participation in Program Activity		.30-.93		.72	18	.56		.77		
SIS total		.69-.91		.86	18	.91		.81		
Observation situations										
One-to-One Interview						.94		.74		
Mealtime					15	.88		.66		
Unstructured Group					14	.81		.86		
Structured Group Activity					16	.78		.90		
Structured Verbal Group					8	.92		.81		

Note. Portions of this table are from *Bay Area Functional Performance Evaluation* (2nd ed.) (p. 15) by S. L. Williams and J. Bloomer, 1978 and 1987, Palo Alto, CA: Consulting Psychologists Press. Copyright 1978 and 1987 by Judith Bloomer and Susan Williams. Adapted by permission. Further reproduction is prohibited without the publisher's consent. SIS = Social Interaction Scale.

^a *n* shown where number of cases used is less than the full sample.

^b Field study, six groups. Correlations shown are range and unweighted averages of the six group correlations.

^{*} Pearson product-moment correlations with two-tailed *p* values. All *p* values are ≤.001 except where otherwise indicated.

are less circumscribed than those of the TOA. Note that Group 3 parameter correlations were higher but more variable than those for Group 4. Group 3 situation correlations were also higher than those of Group 4 but less variable.

The second objective of the study was to determine whether the substitution of the five specific observation situations for the single global rating of the original SIS improved its reliability. Table 5 compares the ranges and averaged correlations for both the original and revised SIS and shows that all parameter correlations were equaled or increased with the revised edition.

The third objective of the study was to assess the internal reliability of the revised SIS. As with the TOA, the SIS data were analyzed for an initial assessment of reliability with a Pearson product-moment correlation procedure. For this analysis, only Rater 1 data were used, which made it possible to include data from Group 1. Correlation matrices showed that intercorrelations among the seven parameters were generally higher than those of the observation situations ($r = .38$ to $.87$ compared with $r = .35$ to $.73$ —see Williams & Bloomer, 1987, for complete information). Among the parameters, Social Behavior correlated most highly with the others ($.65$ to $.87$) and Relations with Authority Figures correlated least ($.38$ to $.69$). Among the observation situations, the Mealtime situation correlated most highly with the other four situations ($.59$ to $.73$) and the Structured Verbal Group situation had the lowest correlations ($.35$ to $.59$).

These data support the contention that the parameters are related components of the social interaction construct, but also that they assess different aspects of social interaction. The existence of generally modest correlations among the observation situations emphasizes that each of the five social situations contributes independent information and should be retained.

The last objective of this part of the study was to determine the degree to which the scores on the new patient Self-Report of Social Interaction correlated with the raters' scores. In short, the correlations were very low and only one parameter, Ability to Work With Others, showed correlations over $.40$. There was an indication that some patients had difficulty understanding the rating guidelines and required assistance to complete the form. Because of the low correlation level it was decided not to include these scores in data used to calculate the SIS parameter scores for this study, and not to include the Self-Report as a scored part of the SIS. However, verbal reports from the raters indicated that the Self-Report was useful in assessing the patient's insight into his or her own social behavior. Therefore, the Self-Report is included as an

optional part of the revised SIS for use at the therapist's discretion.

In summary, although one must consider that the reliability of the results might possibly be limited by incomplete data, the interrater reliability for the revised SIS appears to be higher than for the original. The internal reliability data support the continued use of all seven parameters and all five observation settings. The problem encountered with so many missing two-rater data underscores the difficulty of collecting such research data in the course of ongoing clinical work, as well as the importance of better training in the research procedures.

Validity of the BaFPE

In their original research, the authors completed a number of data analyses to test the validity of the BaFPE as a measure of functional performance. One major analysis was a concurrent validity study comparing BaFPE scores with those of two other measures of functioning, the Global Assessment Scale (GAS) (Endicott, Spitzer, Fleiss, & Cohen, 1976) and a modified version of the Functional Life Scale (FLS) (Sarno, Sarno, & Levita, 1973; Underwood, 1978). The composite BaFPE scores correlated with the GAS at $r = .57$ and with the FLS at $r = .58$, both significant at the $.001$ level (for complete results, see Bloomer & Williams, 1979, p. 94). These results can be considered quite good for a comparison with two different instruments purporting to measure the same area of concern.

In addition, mean scores on the BaFPE for 62 patients were compared with mean scores for 20 normal subjects during the original research, and in the later field study of the original BaFPE, data from the seven subject samples were pooled to compare mean scores of 51 patients and 50 normal subjects. The results of these analyses are shown in Table 6. Differences between patients and normal subjects were found to be significantly different in both studies, although the differences were not so pronounced in the field study. No normal subjects were assessed with the revised BaFPE; therefore, no comparable analysis was done.

A comparison of scores of patients assessed at either admission or discharge showed definite differences between patients in the original study, but the differences did not achieve significance. A comparison of pooled data for the patient subjects in the field study showed relatively little difference between patients at admission and at discharge. In the reliability study of the revised BaFPE, patient subjects were assessed only at admission, hence, no comparable analysis was possible. Thus, it is not yet clear whether the BaFPE can differentiate the functional levels of patients at the beginning and end of treatment in a short-term acute psychiatric setting. It has been diffi-

Table 6
Comparison of Mean Scores of Patient and Normal Subjects in Original BaFPE Research and Field Study

	Original Study					Field Study				
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
TOA										
Normal	20	230.9	3.99			50	226.1	9.92		
Patient	62	187.5	36.71	9.09	.001	51	199.6	33.54	5.39	.001
SIS ^a										
Normal	20	33.1	1.51			50	33.0	2.39		
Patient	60	25.5	5.47	9.43	.001	44	26.1	4.01	10.47	.001
BaFPE ^b										
Normal	20	462.9	9.58			50	457.4	22.53		
Patient	60	365.0	65.55	11.11	.001	44	382.1	55.45	8.92	.001

Note. BaFPE = Bay Area Functional Performance Evaluation. TOA = Task-Oriented Assessment. SIS = Social Interaction Scale.

^a Analysis based on unweighted SIS scores.

^b Composite BaFPE included weighted SIS scores.

cult to test patients at discharge, and more data are needed to provide more definitive answers.

A number of analyses were done with both the original and revised BaFPE instruments to explore whether or not the BaFPE would show different patterns of scores for patients with different diagnoses. Results of the original research showed some possible relationships with diagnosis, but this was not confirmed in the field study. Williams and Houston (1986) reported on an exploratory study of scores on the revised TOA and SIS for patients with different diagnoses. Although depressed patients tended to score somewhat higher than the other diagnostic groups, the differences did not appear to be significant. Again, further studies are needed with larger numbers of patients.

Although the authors have not been able to conduct further formal validity studies, a number of validity studies have been done by other clinicians in the United States and Canada that have contributed to the body of information on the BaFPE as a measure of functional performance. Some of these augment the initial validity studies reported in the research edition of the BaFPE, and some contribute to the validation of the revised instruments. The authors have heard of many studies using the BaFPE, but only those studies for which reports have been received are reported here. Table 7 summarizes the information about these studies.

Studies Using the Original BaFPE

Cheeseman (1980) investigated the validity of the BaFPE for 20 patients with brain vascular disease at a university hospital and Veterans Administration medical center. She compared BaFPE scores with scores on the Functional Life Scale (FLS) (which also was used in the original BaFPE research) and on the Jebsen Hand Function Test. Total BaFPE and FLS scores showed a correlation of .62, supporting the original BaFPE research results. On the Jebsen Hand Function

Test, the skill level of the nonaffected hand of the stroke patient correlated with BaFPE scores at the .0005 significance level, which suggests that adequate motor skills are necessary to successfully perform tasks on the TOA.

Kaufman (1982) compared the Kohlman Evaluation of Living Skills (KELS) with BaFPE scores for 16 psychiatric inpatients with diagnoses of schizophrenia or major affective disorders. She found a significant correlation between scores on the KELS and the BaFPE (total BaFPE, .84; TOA, .84; and SIS, .74) and all correlations were significant at the .001 level. Of the five TOA tasks, the House Floor Plan correlated most highly with the KELS (.82), although three of the five tasks and 7 of the 10 functional parameters correlated with the KELS at $r = .64$ to $.80$, at the .01 level of significance. Descriptive statistics suggested a correlation between the high or low functional range of the BaFPE and KELS scores, indicating higher or lower potential for independent living, but there were not sufficient data for valid statistical analysis.

Bortone (1984) compared BaFPE and GAS scores for 30 schizophrenic and borderline personality patients. She found that a higher composite BaFPE score correlated positively with a diagnosis of borderline personality and that a lower BaFPE score correlated positively with schizophrenic disorders. Borderline patients showed lower scores on the TOA than on the SIS, whereas the opposite was true for schizophrenic patients. Only 2 of 10 TOA parameters correlated significantly with diagnosis, compared with 4 of the SIS parameters. Composite BaFPE scores correlated positively with the GAS but did not reach significance, perhaps because of multiple GAS raters and a small GAS sample of 23 patients. Lastly, an "interrater agreement analysis" (p. 51) using 6 of the TOA parameters and all of the SIS parameters showed agreement ranging from moderate to perfect, with most in the substantial or almost-perfect range.

Accardi (1985) compared the results of three

Table 7
Summary of Studies Using the BaFPE

Name	Date	Version Used	Number of Subjects	Type of Study	Results
Cheeseman	1980	Original	20 CVA patients	Validity	High correlation between BaFPE/FLS. High correlation between TOA/unaffected hand of CVA patient.
Kaufman	1982	Original	16 acute psychiatric patients	Validity	High correlation between BaFPE/KELS.
Bortone	1984	Original	30 schizophrenic borderline inpatients	Reliability and validity	High interrater reliability for SIS. High BaFPE scores for borderline, low for schizophrenic. SIS parameters more significantly related to diagnosis than TOA.
Accardi	1985	Original	19 psychiatric inpatients	Validity	High correlation between three evaluations: BaFPE, NOSIE-30, adaptation of CEBLS.
Brockett	1985	Original	50 psychiatric inpatients	Validity	Significant difference between patient scores in San Francisco and Vancouver.
Francis	1986	Original	20 schizophrenic patients & 20 normal subjects	Validity	Modified version of two tasks did not significantly affect score.
Olson and Jamal	1987	Original	211 psychiatric inpatients	Validity	Developed system of percentiles for reporting BaFPE scores. High correlation between BaFPE scores and community placement. High correlation between BaFPE scores/recommendation for conservator. No difference on BaFPE scores for diagnostic categories.
Mason	1985	Revised	18 psychiatric inpatients	Validity	Low correlation between self-report and observer's ratings on SIS. Low correlation between TOA parameters and SIS parameters.
Thibeault and Blackmer	1987	Revised	60 schizophrenic & depressed patients	Validity	Difference between groups on only two parameters of TOA. ECT/age were best predictors of BaFPE scores.
Newman	1987	Revised	21 psychiatric inpatients	Validity	High correlation between ACL, BaFPE, GAS. High correlation between TOA tasks and ACL, except Sorting Shells. ACL correlation highest with Affective component of TOA.

Note. TOA = Task-Oriented Assessment. CVA = cerebrovascular accident. BaFPE = Bay Area Functional Performance Evaluation. FLS = Functional Life Scale. KELS = Kolman Evaluation of Living Skills. SIS = Social Interaction Scale. NOSIE-30 = Nurse's Observation Scale for Inpatient Evaluation. CEBLS = Comprehensive Evaluation of Basic Living Skills. ECT = electroconvulsive therapy. ACL = Allen Cognitive Level Test. GAS = Global Assessment Scale.

evaluations of 19 psychiatric patients. The evaluations used were the BaFPE, the Nurse's Observation Scale for Inpatient Evaluation (NOSIE-30), and an adaptation of the Comprehensive Evaluation of Basic Living Skills (CEBLS). Scores on the TOA correlated with total NOSIE-30 scores at .66 ($p \leq .001$), with the adapted CEBLS at .63 ($p \leq .005$), and with one NOSIE-30 subtest, Social Competence, at .86 ($p \leq .001$). SIS scores correlated with NOSIE-30 scores at .69 ($p \leq .005$) and with three NOSIE-30 subtests and the adapted CEBLS at .46 to .56 ($p \leq .05$). Accardi's findings support the use of the BaFPE as a predictor of functional performance, as determined by two other functional assessments given concurrently.

Of special interest to therapists in Canada was the potential cultural bias of the BaFPE. Brockett (1985)

reported that therapists at Vancouver General Hospital tested 50 patients in the psychiatric unit and found significant differences between scores of the San Francisco patients and those of the British Columbia patients. Brockett hypothesized that cultural differences may have contributed to this finding and suggested that some TOA tasks, such as the Bank Deposit Slip, were heavily dependent on life experiences. This study underscores the importance of developing extended norms, but reminds us that local norms may be needed for evaluating a specific population.

Francis (1986) examined two TOA tasks to see if modification of the tasks would make the testing medium more appropriate for patients in her mental health center. She substituted buttons for shells in the Sorting Shells task, and changed the sample floor plan and the room list in the House Floor Plan task. Testing

20 schizophrenic and 20 normal subjects, she administered the original tasks to half of each group and the modified tasks to the other half. On both tasks, the schizophrenic patients scored lower than the normal subjects, as predicted. However, the modified versions of the tasks did not differentially influence task performance for either group. (The revised BaFPE likewise uses an altered sample floor plan and list of rooms, but the alterations are different from those proposed by Francis.)

Olson and Jamal (1987) examined several clinical applications of the BaFPE in an extensive study involving 211 psychiatric inpatients. They developed percentile scores that could be used to interpret BaFPE scores to interdisciplinary treatment teams. In their use of the BaFPE they found that scores on the TOA correlated significantly with the placement of patients in three different living situations after hospitalization: Patients with the lowest scores were placed in board-and-care or semicustodial living situations; those with midrange scores were placed in continuing treatment situations; and those with the highest scores were discharged to their families or to independent living situations. Scores on the SIS did not correlate with those placements. They also made recommendations partially based on the BaFPE scores regarding the need for legal conservatorship for patients. Actual treatment team decisions regarding conservatorship correlated significantly with BaFPE-based recommendations.

In addition, they computed internal correlations for three main scales of the BaFPE (TOA, SIS, and the composite BaFPE), and all these correlated highly. Lastly, they looked at scores on the three main scales in relation to diagnoses of schizophrenia, depression, manic-depression, and others. They found no significant difference for any of the diagnoses, thereby supporting the findings of Thibeault and Blackmer (1987) reported below.

Studies Using the Revised BaFPE

A Self-report of Social Interaction was initially included in the draft revision of the BaFPE. Mason (1985) compared this self-report for 18 inpatients to observer ratings on the SIS and found a positive correlation with only one SIS parameter, Verbal Communication. These findings suggested that the validity of the self-report as a formal scored part of the SIS was questionable and contributed to the BaFPE authors' decision not to include it in the scoring of the SIS.

Secondarily, Mason's study supported the continued use of both the TOA and the SIS, because most functional parameters of the TOA did not correlate significantly with those on the SIS, which suggests that they are in fact measuring two different aspects of functional performance.

Newman (1987) examined the relationship between cognitive level and performance on the TOA. She determined the cognitive level of 21 inpatients by administering the revised Allen Cognitive Level Test (ACL). A significant correlation was found between the revised ACL and the TOA ($.63 \leq .01$), as scores on the TOA increased with the cognitive level. In addition, Newman found that both the TOA and the ACL correlated significantly with the GAS. In the TOA tasks, all except the Sorting Shells correlated significantly with the ACL at the .05 level, with the Block Design showing the highest correlation (.65). In an unexpected finding, the ACL showed the highest correlation with the Affective component of the TOA, rather than with the Cognitive or Performance components, as expected.

A validity study investigating the relationship between TOA scores and diagnosis was reported by Thibeault and Blackmer (1987). They compared the revised TOA with three subscales of the Weschler Adult Intelligence Scale (WAIS) for 60 psychiatric inpatients. The three WAIS subscales all correlated significantly with the TOA (.67, .58, and .60, $p \leq .001$). They also compared scores for two diagnostic groups, schizophrenia and depression, and found that these differed significantly on only one of the TOA parameters, Frustration Tolerance, and on the task-specific qualitative observations, which suggests that diagnostic groups may not show many significant differences on the TOA. There were significant differences between the two groups on the Qualitative Signs and Referral Indicators section, but this section is not formally scored on the TOA. In addition, they studied the relationship between the TOA and several demographic and treatment factors. They found no significant association with level of medication, sex, length of stay in the hospital, or number of admissions. However, they found that age and level of education correlated significantly with TOA scores, with older and less educated patients scoring lower. Patients who had received electroconvulsive therapy (ECT) scored significantly lower on both the TOA Block Design and the WAIS Block Design and on 10 of the 12 TOA parameters than did patients who had not received ECT. Thibeault and Blackmer suggested that norms may need to be developed that take age and education into consideration.

Conclusions

The results of the interrater reliability analyses for both the TOA and the SIS indicated that the revised instruments were more reliable than the originals. In addition, the therapists who collected data for the study indicated that the revised instructions were easier to understand and the revised instruments were

easier to use, which resolved criticism leveled at the original instruments. However, the results also continued to show more than desirable variability on some parameters of both instruments. In addition, there was more variability within each group's data than desired. This variability tends to support the need for providing formal training to users to increase the reliability of their use of the instruments. Such training was not provided in any of the studies we conducted because we wanted to find out how reliably the instruments could be administered by using the manual as the primary mode of instruction. Our experience seems to indicate that an increase in interrater reliability would require that training methods, along with a method of evaluating such training, be developed.

Study of the SIS reliability was hampered by limited two-rater data reported above. Only two of four teams provided enough two-rater data to support valid data analysis. All the teams had difficulty finding time in their programs to jointly rate the patients in the four group social settings required for completion of the SIS—particularly since the therapists usually worked with different programs and patients. There was no clear feedback regarding the usefulness of the multisite format of the revised SIS and its relative ease of use by a single therapist (as opposed to a pair of therapists). Further studies are needed to provide more information about both the reliability of the SIS and the usefulness of the multisite ratings as a basis for the SIS score.

National norms are not yet available to compare with the performance of individual patients on the BaFPE, and this makes it difficult to interpret results of testing with the BaFPE instruments. Until normative data are available, it is suggested that interpretation be based on the preliminary norms presented in the revised BaFPE manual and on the clinical judgment of the therapist. Populations evaluated so far in the BaFPE studies we initiated have been inpatients treated in acute psychiatric hospitals or available groups of normal subjects, including hospital staff and trainees, friends and relatives. It has been suggested that the BaFPE instruments might be useful with other patient populations (e.g., cerebrovascular accident patients, retarded adolescents), but reliability, validity, and norms for such patient groups remain to be established.

Feedback indicates that some therapists who like the BaFPE instruments do not use them "because they take too much time for an acute psychiatric program." Others have indicated that they have developed their own short forms of the TOA (the selection of tasks varies by the therapist) or have adapted the TOA rating system to other "more relevant" tasks or to a group activity setting. We are interested in seeing the

development of a short form of the TOA and, possibly, of the SIS, but urge interested therapists to go about this process using valid research approaches. Present standardization results do not apply to these short ad hoc forms, only to the complete instruments. The internal reliability data reported above provide indications of possible valid short forms, but much more study is needed, including testing of short form results with results of the complete instruments. It is hoped that graduate students and therapists will consider undertaking such studies.

The determination of the validity of an evaluation instrument is a lengthy process requiring many studies by many investigators in many different settings over time (Benson & Clark, 1982). The studies done by other clinicians that are reported above have augmented the reliability and validity studies we have done. They have shown significant positive correlations between the BaFPE and several other measures of patient functioning. However, there is an urgent need for many more data about the relationship of BaFPE scores to specific living skills, such as are provided by Kaufman's (1982) comparison with the KELS. One study provided support for the original BaFPE's ability to predict a patient's placement after discharge from the hospital, but no formal predictive studies have yet been conducted with the revised BaFPE and these are needed. Above all, many more data need to be reliably collected about all kinds of psychiatric patients and normal people in different parts of the United States and in other countries, so that appropriate norms can be developed. At present, a major effort is under way at the State University of New York at Buffalo to collect both patient and normal subject data for the TOA.

At this time, the important initial stages of development and revision of the BaFPE instruments have been completed, and the more reliable instruments are available for both clinical and research use.

Acknowledgments

We are indebted to the clinicians who have assisted with data collection and have given us feedback since we began work on the BaFPE in 1978. We are also extremely grateful to the following organizations, which have provided funding support at different stages of the development of the BaFPE: (a) the American Occupational Therapy Foundation; (b) the California Foundation for Occupational Therapy; (c) the Occupational Therapy Association of California, Golden Gate Chapter; (d) the Occupational Therapy Association of California, Santa Clara Chapter; and (e) the Langley Porter Psychiatric Institute, University of California, San Francisco (Biomedical Research Support Grant NIH No. RR-05755-05).

References

Accardi, M. (1985). *The Bay Area Functional Performance Evaluation: A validity study*. Unpublished master's

thesis, Boston School of Occupational Therapy, Tufts University, Boston, MA.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* [DSM-III] (3rd ed.). Washington, DC: Author.

Barris, R., Kielhofner, G., & Watts, J. (1983). *Psychosocial occupational therapy: Practice in a pluralistic arena*. Laurel, MD.: RAMSCO.

Benson, J., & Clark, F. (1982). A guide for instrument development and validation. *American Journal of Occupational Therapy*, 36, 789-800.

Bloomer, J. S., & Williams, S. K. (1979). *The Bay Area Functional Performance Evaluation* (research ed.). Palo Alto, CA: Consulting Psychologists Press.

Boring, E. G. (1957). *History of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

Bortone, J. (1984). *Functional component skills associated with DSM-III diagnoses of schizophrenia and borderline personality disorders*. Unpublished master's thesis, Programs in Occupational Therapy, New York University, New York, NY.

Brockett, M. (1985). *Cultural variations in Bay Area Functional Performance Evaluation scores—Considerations for occupational therapy*. Vancouver, British Columbia: Vancouver General Hospital.

Cheeseman, J. W. (1980). *An investigation of the concurrent validity of the Bay Area Functional Performance Evaluation for patients with brain vascular disease*. Unpublished master's thesis, Virginia Commonwealth University, Richmond, VA.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale. *Archives of General Psychiatry*, 33, 766-771.

Francis, E. (1986). *Comparison of two subtests of the Bay Area Functional Performance Evaluation*. Unpublished master's thesis, Sargent College of Allied Health Professions, Boston University, Boston, MA.

Herrnstein, R., & Boring, F. (1965). *A source book in the history of psychology*. Cambridge, MA: Harvard University Press.

Kaufman, L. (1982). *A comparison of performance on the Bay Area Functional Performance Evaluation and the Hohlman Evaluation of Living Skills by adult psychiatric patients*. Unpublished master's thesis, College of Health Related Professions, University of Florida, Gainesville, FL.

Kielhofner, G. (Ed.). (1985). *A model of human occupation: Theory and application*. Baltimore: Williams & Wilkins.

Kielhofner, G., & Burke, J. P. (1980). A model of human occupation: Part 1. Conceptual framework and content. *American Journal of Occupational Therapy*, 34, 572-581.

Kielhofner, G., Burke, J. P., & Igi, C. H. (1980). A model of human occupation: Part 4. Assessment and intervention. *American Journal of Occupational Therapy*, 34, 777-788.

Mason, J. (1985). *Observer ratings vs. self-report of social interaction as assessed by the Bay Area Functional Performance Evaluation*. Undergraduate research project, School of Occupational Therapy, Dalhousie University, Halifax, Nova Scotia.

Moorhead, L. (1969). The occupational history. *American Journal of Occupational Therapy*, 23, 329-334.

Mosey, A. C. (1970). *Three frames of reference for mental health*. Thorofare, NJ: Slack.

Newman, M. (1987). *Cognitive disability and functional performance in individuals with chronic schizophrenic disorders*. Unpublished master's thesis, University of Southern California, Los Angeles.

Olson, B., & Jamal, J. (1987). *The BaFPE: Standardization and clinical application in acute adult psychiatry*. Unpublished manuscript, University of California Irvine Medical Center, Irvine, CA.

Reed, K. L. (1984). *Models of practice in occupational therapy*. Baltimore: Williams & Wilkins.

Reilly, M. (1966). A psychiatric occupational therapy program as a teaching model. *American Journal of Occupational Therapy*, 20, 61-67.

Robertson, S. C. (1985). Generic frames of reference in occupational therapy. In M. Kirkland & S. C. Robertson (Eds.), *PIVOT: Planning and implementing vocational readiness in occupational therapy* (pp. 133-136). Rockville, MD: American Occupational Therapy Association.

Sarno, J. E., Sarno, M. T., & Levita, E. (1973). The Functional Life Scale. *Archives of Physical Medicine and Rehabilitation*, 54, 214-220.

Thibeault, R., & Blackmer, E. (1987). Validating a test of functional performance with psychiatric patients. *American Journal of Occupational Therapy*, 41, 515-521.

Underwood, P. (1978). *Nursing care as a determinant in the development of self-care behavior by hospitalized adult schizophrenics*. Unpublished doctoral dissertation, University of California, San Francisco.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158-177.

Williams, S. L., & Bloomer, J. (1987). *Bay Area Functional Performance Evaluation* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Williams, S. L., & Houston, D. (1986). Use of the Bay Area Functional Performance Evaluation (BaFPE) with the depressed patient: A preliminary impression. In *Depression: Assessment and treatment update: Proceedings* (pp. 22-35). Rockville, MD: American Occupational Therapy Association.