ABSTRACT
                  The development and validation of the General
Management In-Basket (GMIB) is described. The GMIB is a theory-based
generic in-basket simulation, designed to assess supervisory and
management skills independent of any job classification. Three of the
15 in-basket items in the GMIB are critical and are scored on a 0-5
scale. The remaining 12 items are scored on a 0-4 scale. Using an
item-by-item scoring format, total mean score interrater
reliabilities across 4 studies have ranged from 0.86 to 0.95. The
GMIB requires only 20 minutes to scores and needs only one rater. In
a concurrent validation of the GMIB involving incumbents from 120 job
classifications, an estimated true validity of 0.41 was found for
predicting the composite of ratings by immediate supervisors on 6 job
performance dimensions (n=275). An estimated true validity of 0.44
was found for predicting the composite of ratings by second-level
supervisors on the same performance dimensions (n=239). Factor
analysis identified four factors for the GMIB, making it possible to
retain the dimension format associated with traditional in-baskets.
Automated candidate feedback can also be generated with the GMIB.
Twelve tables present data from the studies, and an 11-item list of
references is included. (SLD)

TRADITIONAL IN-BASKETS

VS.

THE GENERAL MANAGEMENT IN-BASKET (GMIB)

presented at the
International Personnel Management Association
Assessment Council Conference
June 23-27, 1991

by:

Richard C. Joines, President, Management & Personnel Systems, Inc.

1475 North Broadway * Suite 280 * Walnut Creek, California 94596 * 415-932-0203

2

THE GENERAL MANAGEMENT IN-BASKET

Overview

This paper describes the development and validation of the General Management In-Basket (GMIB). The GMIB is a theory-based, generic in-basket, designed to assess supervisory/managerial skills independent of any particular job classification. The General Management In-Basket departs from traditional dimension scoring approaches which are complex, time consuming and often plagued by unreliability. By using an item-by-item scoring format, total mean score inter-rater reliabilities across four studies have ranged from .86 to .95. The time required to score the GMIB is less than 30 minutes and requires only one rater due to the high single rater reliabilities obtained in scoring the GMIB.

In a concurrent validation of the GMIB involving incumbents drawn from 120 job classifications and ranging from professional staff up through the third level of supervision, an estimated true validity of .41 was found for predicting the composite of ratings by immediate supervisors on six job performance dimensions (N = 275, p<.01). An estimated true validity of .44 was found for predicting the composite of ratings by second-level supervisors on the same six performance dimensions (N = 239, p<.01).

Although the GMIB is scored on an item-by-item basis as opposed to the traditional dimension scoring format, factor analysis identified four factors; thus, it has been possible to retain the "dimension" information associated with traditional in-baskets. Automated candidate feedback reports by dimension can be generated as a result of developing a large bank of descriptive narrative statements covering the types of responses that have been identified at each score level on each of the 15 GMIB items.

KEY PROBLEMS WITH TRADITIONAL IN-BASKET TECHNOLOGY

The time required to score traditional in-baskets is two to five hours and this may account for the paucity of research on in-baskets (Thornton & Byham, 1982). On a practical level, many organizations may be hesitant to use an in-basket when examining a large number of candidates due to high scoring time (Kraus, 1986).

In a recent review of the reliability and validity of in-baskets, (Schippmann, et. al., 1990), the authors conclude that in-baskets can be reliably scored but that obtained reliability coefficients are "modest at best". With regard to inter-rater reliability, the authors included 13 studies in their review, with 10 studies reporting the obtained range of coefficients across rated dimensions and three studies reporting a single reliability coefficient for in-basket performance.

In an effort to obtain some rough quantification of the inter-rater reliability studies reviewed by Schippmann, et. al., the mean of the coefficients forming the low end of the range was calculated for the 10 studies reporting a range of reliabilities. The resultant mean reliability was .60. For the three studies reporting a single reliability coefficient, the mean reliability was found to be .77. Thus, the authors characterization of obtained reliability coefficients as "modest" appears warranted.

With regard to in-basket validity, Schippmann et. al. conclude that validity "is at best marginal and generally higher in settings where the in-basket was specifically constructed for a defined target job". The studies included a wide-range of criterion measures, from indirect measures such as grades, scores on standardized exams and salary progress to more direct measures such as supervisory ratings of on-the-job performance. For the more direct measures of performance, 13 studies report a range of validity coefficients for the in-basket dimensions.

- 2 -

The means of the validity coefficients forming the low and high ends of the reported ranges were calculated for the 13 studies in this category, resulting in a mean range of validity coefficients of -.25 to .36. The mean negative validity forming the low end of this range was not due to any single large negative validity coefficient, but rather, a host of negative values.

Based on this global analysis, the characterization of in-basket validity as being "marginal" also seems to be warranted. In short, the available research on in-baskets suggests that they are plagued by problems related to reliability and that these problems are not offset by high validity.

The traditional approach to scoring in-baskets is difficult and complex which no doubt is a major reason for the disappointing reliability results. One of the major reasons for these results may be the lack of standardization of the behavioral information base. Candidates may choose to complete different items in the time allotted and the number of items completed by candidates may vary significantly. These differences create considerable difficulties for the raters in assigning dimension scores.

In addition to the fact that the items completed by the candidates may differ significantly, there is a related standardization problem even for those items that all candidates may complete. This problem stems from the fact that candidates are totally free to respond as they choose on each item. Since candidate responses may vary greatly, the same item will not necessarily provide evaluative information on the same assessment dimensions.

For example, one candidate might choose to write a memo in response to an in-basket item. The completed memo might require considerable interpersonal sensitivity in order to be considered appropriate to the situation. Another candidate, however, might choose to delay action on the item. For the first candidate, the item elicits information relative to the dimensions written communications and interpersonal sensitivity, whereas for the second candidate, information on these dimensions is not elicited by the item.

Given many such combinations across even a small number of in-baskets, raters find that they must rate dimensions such as problem analysis, judgment, management control, interpersonal sensitivity, etc., based on a different behavioral information base for each candidate. As the number of candidates increases, raters find it increasingly difficult to insure that they have rated all in-baskets using the same rating standards. This may lead to more and more time devoted to the review of previously rated in-baskets in an effort to be consistent and fair.

The General Management In-Basket (GMIB) Scenario and Item Types

Candidates taking the GMIB assume the role of the Director of the XYZ Division of a hypothetical organization (either public sector or private sector). There are 15 items that deal with the following kinds of general supervisory/management issues:

* Employee motivation & morale
* Managing organizational conflict
* Implementation of new procedures
* Dealing with personnel external to the organization
* Delegation
* Performance problems
* Staff development/growth
* Work organization/efficiency
* Group dynamics/team efforts

- 3 -

Candidate Response Format

For each item attempted, candidates respond on standardized forms, of which there are two. On the first, candidates are instructed to analyze the supervisory/managerial issues involved in the item -- even if they believe the item requires no immediate action. Subsequently, and on the same form, candidates are instructed to describe any actions they would take in handling the item, either in the present or at a future date. Finally, a second form is supplied to candidates on which they are to write any memos or letters that they would write in handling the in-basket item. Each form is numbered to correspond to the item. Only the response forms are necessary to score the in-basket since all analyses and actions must be shown on these forms.

Scoring Method

The GMIB is scored on an item-by-item basis. Detailed rating guidance exists for each item. This guidance includes a narrative discussion explaining the goal of the item and an analysis of the management issues that are involved. In addition, rating scales anchored with descriptions of the responses to be rated at each level, are used. Three of the 15 in-basket items are considered critical and these are scored on a scale ranging from 0 to 5. The remaining 12 items are scored on a scale ranging from 0 to 4.

Basis for Scoring Guidance

In an effort to devise an in-basket that would have validity across a wide range of management situations, a number of the items are geared toward the application of management theory to practice. In particular, concepts related to McGregor's Theory Y, participative management, and situational leadership are utilized in the narrative scoring guidance. Since some items have this orientation, the proper handling of the items is based more on acceptance of the theory and of its application to an applied situation than on the judgments of particular subject matter experts or assessors. Thus, the scoring guidance, as currently formulated, would not be valid in an organization that desired an autocratic style of leadership.

Other items in the in-basket are based on commonly accepted principles of organizational effectiveness and sound management practice. Items in these categories include issues such as dealing with a performance problem to responding to an important public official on a sensitive matter.

Reliability of Ratings

Four inter-rater reliability studies have been conducted on the GMIB. The range of obtained single rater reliability coefficients on total in-basket score is .86 - .95. These results are summarized in Table 1.

With regard to the reliability of item ratings, study #4, which included the largest number of in-baskets, found a mean item inter-rater reliability of .87.

Where inter-rater reliability is high, there are decreasing returns associated with adding additional raters. The Spearman-Brown formula may be used to estimate the reliability of the scoring process using two raters instead of one. Given a single rater reliability of .86, using two raters increases the reliability to .92. If the single rater reliability is .95, the reliability with two raters is .97. Thus, the improvement is negligible in both cases and does not warrant adding a second rater. Clearly, this has significant implications for savings in terms of time and costs.

5

TABLE 1

INTER-RATER RELIABILITY FOR GMIB TOTAL SCORE

	No. Raters	# In-Baskets	Mean r
Study #1	6	10	.87
Study #2	5	10	.94
Study #3	4	10	.93
Study #4	2	100	.95

Time Required to Score

Trained raters average approximately 20 minutes to thoroughly score a test. This includes assigning scores on each item attempted by the candidate as well as selecting from among a bank of narrative statements descriptive of the candidate's performance on each item.

When the rater has completed the scoring form, the data is entered into a custom data base program, requiring approximately one minute per candidate. The data base program then sends a file to a word processor which then automatically produces a 5 - 6 page narrative report on the candidate.

This obviously compares quite favorably with traditiona, approaches to scoring and developing candidate feedback reports.

Criterion-Related Validation

Three hundred sixty-five employees of a public sector organization completed the GMIB as an initial hurdle in competition for selection into an advanced management development program. The sample consisted of incumbents in levels 2 - 4 of the organization's classification structure. Employees from approximately 120 separate job classifications were represented in the sample.

Level two applicants were non-supervisory higher level professional personnel. Level three applicants weru generally first level supervisors and level four applicants were either second or third level supervisors. There were 219 level two candidates, 102 level three candidates and 44 level four candidates.

Performance ratings were concurrently collected from immediate and ne..t higher level supervisors. The number of completed performance rating forms was 278 for immediate supervisors and 243 for next-higher-level supervisors. Ratings on 194 by both raters were available. This permitted an evaluation of the reliability of the criterion measures as well as the formation of several overall composite measures based on both sets of ratings.

Ratings were made on a nine point rating scale (1 = low; 9 = high) on the following performance dimensions:

(1) written communication skill; (2) leadership skill; (3) interpersonal relations skill; (4) planning and organizing skill; (5) analyzing problems and making sound decisions; and (6) oral communications skill.

For each performance dimension, two ratings were made: (1) the employee was rated in relation to employees at the "same" organizational level; and (2) the employee was rated in relation to "all" employees at organizational levels 2, 3 and 4. After rating employees on the performance dimensions, raters were asked to supply an overall rating of the employee in relation to employees at the same organization level. The same nine point rating scale was used. The mean of the overall ratings supplied by immediate and next-higher-level supervisors served as a subjective overall composite criterion measure.

In addition to the subjective overall ratings and composite measure, a series of mechanically derived overall measures of performance were formed, as follows: (1) mean of immediate supervisor's ratings on the six performance dimension ratings using candidates at the "same" organizational level as the reference group; (2) mean of next-higher-level supervisor's ratings for the "same" level reference group; (3) mean of immediate supervisor's ratings on the six performance dimensions using "all" eligible candidates as the reference group; (4) mean of next-higher-level supervisor's ratings on the six performance dimensions using "all" eligible candidates as the reference group.

Estimates of the reliability of the in-basket and the performance dimension ratings made by supervisors were obtained using Cronbach's (1951) coefficient alpha based on 99 cases for which complete data on all in-basket items and performance measures was available. These results are given in Table 2.

Table 3 presents in-basket item correlations with total in-basket scores. These correlations (n = 365) ranged from .37 to .52.

TABLE 2

COEFFICIENT ALPHA FOR PREDICTOR AND CRITERION MEASURES

Measures	Coefficient Alpha
In-Basket (15 items; Mean = 18.14, SD = 8.44, N = 365)	.71
Criteria	
"Same" Level Ratings by Immediate Supervisors	.92
"All" Levels Ratings by Immediate Supervisors	.92
"Same" Level Ratings by Next-Higher-Level Supervisors	.91
"All" Levels Ratings by Next-Higher-Level Supervisors	.92

- 6 -

7

TABLE 3

IN-BASKET ITEM - TOTAL CORRELATIONS (N = 365)

Items	Pearson r
1	.51
2	.37
3	.47
4	.48
5	.42
6	.50
7	.46
8	.37
9	.39
10	.42
11	.41
12	.52
13	.43
14	.48
15	.46

Validity Coefficients

Table 4 gives the obtained validity coefficients for total in-basket score in relation to the subjective overall ratings of performance and the mechanically derived overall measures. As will be noted, all validity coefficients were highly significant.

Table 5 gives the obtained validity coefficients for the six performance dimensions based on ratings by immediate and next-higher-level supervisors for the "same level" reference group. All validity coefficients are highly significant.

Table 6 gives the obtained validity coefficients for the six performance dimensions based on ratings by immediate and next-higher-level supervisors for the "all levels" reference group. Once again, all validity coefficients are highly significant.

8

TABLE 4

OBTAINED VALIDITY COEFFICIENTS: CORRELATION OF TOTAL IN-BASKET SCORE WITH OVERALL MEASURES OF PERFORMANCE BY IMMEDIATE AND NEXT-HIGHER-LEVEL SUPERVISORS BASED ON "SAME" LEVEL AND "ALL" LEVELS REFERENCE GROUPS

Immediate Supervisors			Next-Higher-Level-Supervisors		
Subjective Overall Rating	Same Level Overall Measure	All Levels Overall Measure	Subjective Overall Rating	Same Level Overall Measure	All Levels Overall Measure
r = .28* (n = 274)	r = .31* (n = 272)	r = .31* (n = 275)	r = .29* (n = 239)	r = .34* (n = 238)	r = .33* (n = 239)

* p < .0001

TABLE 5

OBTAINED VALIDITY COEFFICIENTS: CORRELATION OF TOTAL IN-BASKET SCORE WITH RATINGS MADE BY IMMEDIATE AND NEXT-HIGHER-LEVEL SUPERVISORS ON SIX PERFORMANCE DIMENSIONS FOR "SAME LEVEL" REFERENCE GROUP

	Immediate Supervisors	Next-Higher Level Supervisors
Written Communications	.29* (n=277)	.36* (n=240)
Leadership	.26* (n=274)	.30* (n=241)
Interpersonal Relations	.19* (n=274)	.18* (n=241)
Planning & Organizing	.30* (n=275)	.32* (n=242)
Analyzing Problems/Decisions	.27* (n=273)	.24* (n=242)
Oral Communications	.27* (n=276)	.31* (n=242)

* p < .01

TABLE 6

OBTAINED VALIDITY COEFFICIENTS: CORRELATION OF TOTAL IN-BASKET SCORE WITH RATINGS MADE BY IMMEDIATE AND NEXT-HIGHER-LEVEL SUPERVISORS ON SIX PERFORMANCE DIMENSIONS FOR "SAME LEVEL" REFERENCE GROUP

	Immediate Supervisors		Next-Higher Level Supervisors	
Written Communications	.38*	(n=241)	.38*	(n=241)
Leadership	.23*	(n=242)	.30*	(n=241)
Interpersonal Relations	.19*	(n=243)	.18*	(n=241)
Planning & Organizing	.32*	(n=242)	.32*	(n=242)
Analyzing Problems/Decisions	.28*	(n=243)	.24*	(n=242)
Oral Communications	.31*	(n=243)	.31*	(n=242)

* $p < .01$

In order to obtain an estimate of the reliability of the performance rating criterion measures, the ratings of immediate and next-higher-level supervisors were correlated, as follows: (1) subjective overall ratings of performance; (2) mean of six dimension ratings for the "same level" reference group; (3) mean of six dimension ratings for the "all levels" reference group; (4) ratings on the six performance dimensions for the "same" level reference group; and (5) ratings on the six performance dimensions for the "all" levels reference group.

Reliabilities consistent with published research were found for the subjective overall measures of performance and the "same level" mechanically derived composite measure (.62 and .61, respectively). The reliability of the "all levels" mechanically derived composite was .56. The reliability coefficients for the six performance dimensions ranged from .47 to .61 for the "same" level ratings and from .43 to .51 for the "all" levels ratings. The inter-rater reliability results are given in Table 7.

TABLE 7

OBTAINED INTER-RATER RELIABILITY COEFFICIENTS FOR RATINGS BY IMMEDIATE AND NEXT-HIGHER-LEVEL SUPERVISORS ON SIX PERFORMANCE DIMENSIONS AND OVERALL COMPOSITES

Criteria	Inter-Rater Reliability	
	"Same" Level	"All" Levels
Performance Dimensions		
Written Communications	.51 (n=194)	.46 (n=196)
Leadership	.61 (n=195)	.51 (n=197)
Interpersonal Relations	.47 (n=195)	.45 (n=198)
Planning & Organizing	.52 (r.=196)	.50 (n=197)
Analyzing Problems/Decisions	.50 (n=196)	.50 (n=198)
Oral Communications	.48 (n=196)	.43 (n=198)
Overall Measures of Performance		
Subjective Overall Rating	.62 (n=192)	
Mechanically Derived "Same" Level	.61 (n=192)	
Mechanically Derived "All" Levels	.56 (n=194)	

The obtained validity coefficients were corrected for unreliability in the criterion in order to gain a better estimate of the true validity of selection procedures. Table 8 gives the corrected validity coefficients for the overall measures of performance. These validities are the best estimates available of the true validity of the in-basket in predicting overall performance.

TABLE 8

BEST ESTIMATES OF TRUE VALIDITY: OBTAINED VALIDITIES CORRECTED FOR CRITERION UNRELIABILITY

	Subjective Overall Rating	Mechanical Overall Rating "Same" Level	Mechanical Overall Rating "All" Levels
Immediate Supervisor	.35 (n=274)	.40 (n=272)	.41 (n=275)
Next-Higer Level Supervisor	.37 (n=239)	.44 (n=238)	.44 (n=239)

11

Table 9 gives the corrected validity coefficients for the six performance dimension ratings ("same" and "all" levels). These estimates are the best estimates of the true validity of the in-basket in predicting specific performance dimension ratings by immediate and/or next-higher-level supervisors.

TABLE 9

BEST ESTIMATES OF TRUE VALIDITY: OBTAINED "SAME" AND "ALL" LEVELS DIMENSION VALIDITIES CORRECTED FOR CRITERION UNRELIABILITY

| | "Same" Level | | "All" Levels | |
	Immed.	Next	Immed.	Next
Written Communications	.41	.50	.46	.56
Leadership	.33	.38	.36	.32
Interpersonal Relations	.28	.26	.31	.28
Planning & Organizing	.42	.44	.42	.45
Analyzing Problems/Decisions	.38	.34	.35	.40
Oral Communications	.39	.45	.44	.47

Factor Analysis

An exploratory factor analysis was performed on the scores of the 365 candidates in the sample. The intent was to determine whether independent and interpretable factors could be identified. Therefore, a principal components factor analysis was conducted using a varimax rotation (Kim, 1975). The Kaiser criterion of extracting only factors with an Eigenvalue greater than one was applied.

Four interpretable factors, accounting for 50% of the variance in total scores, were identified and named as shown below:

1. Leadership Style and Practices

2. Handling Priorities and Sensitive Situations

3. Managing Conflict

4. Organizational Practices/Management Control

Factor 1 (Eigenvalue = 3.12) clustered items dealing with an understanding of leadership and motivation principles, along with an understanding of how to vary the amount of direction given subordinates depending on the situation.

Factor 2 (Eigenvalue = 1.86) grouped together those items that represented priority or sensitive public relations matters.

Factor 3 (Eigenvalue = 1.38) clustered those items that involved dealing with existing conflict among staff and/or situations requiring considerable interpersonal skill and insight in order to avoid staff conflict or morale problems.

Factor 4 (Eigenvalue = 1.08) emphasized those items that required an understanding of the importance of organizational goal accomplishment and efficient methods of operation, along with a willingness to redirect subordinate staff as appropriate and necessary.

Two factor scoring methods discussed by Gorsuch (1974, p. 238) were investigated. In method #1, items were allocated to the factor on which they loaded highest. In method #2, all items with salient loadings on a factor (twice the level required for significance) were allocated to the factor. For both methods, rounded weights in half-point intervals were used instead of exact loadings.

Table 10 gives the validity of total factor scores (sum of individual factor scores) for each scoring method in predicting ratings by immediate and next-higher-level supervisors on the overall composite "same" level and "all" levels criterion measures. Table 11 provides estimates of the true validity of each method by correcting for unreliability in the criterion measures.

TABLE 10

OBTAINED VALIDITY COEFFICIENTS: CORRELATION OF SUM OF FACTOR SCORES FOR TWO FACTOR SCORING METHODS WITH OVERALL MEASURES OF PERFORMANCE BASED ON "SAME" AND "ALL" LEVELS RATINGS BY IMMEDIATE AND NEXT-HIGHER-LEVEL SUPERVISORS

	Immediate Supervisors		Next-Higher-Level Supervisors	
	Same Level Overall Measure	All Levels Overall Measure	Same Level Overall Measure	All Levels Overall Measure
METHOD 1	r = .32* (n = 272)	r = .32* (n = 275)	r = .34* (n = 238)	r = .34* (n = 239)
METHOD 2	r = .29* (n = 272)	r = .29* (n = 275)	r = .31* (n = 238)	r = .31* (n = 239)

TABLE 11

BEST ESTIMATES OF TRUE VALIDITY OF SUM OF FACTOR SCORES
FOR TWO SCORING METHODS

	Immediate Supervisors		Next-Higher-Level Supervisors	
	Same Level Overall Measure	All Levels Overall Measure	Same Level Overall Measure	All Levels Overall Measure
METHOD 1	r = .41* (n = 272)	r = .43* (n = 275)	r = .44* (n = 238)	r = .45* (n = 239)
METHOD 2	r = .37* (n = 272)	r = .39* (n = 275)	r = .37* (n = 238)	ı = .41* (n = 239)

ADDITIONAL VALIDITY DATA AND CORRELATIONS WITH OTHER INSTRUMENTS

Clients using the GMIB may also utilize one or more assessment exercises and/or conduct assessment centers. Table 12 summarizes data submitted by client organizations. The sample sizes, while not specified below, typically consist of small groups of candidates, usually 6 - 12. These results indicate a common pattern of substantial correlations with assessment exercises and assessment center results.

TABLE 12

GMIB CORRELATIONS WITH OTHER SELECTION DEVICES

SUPERVISORY LEVEL	SELECTION PROCESS	CORRELATION
3RD LEVEL SUP.	2 DAY A.C.	.55
2ND LEVEL SUP.	DEC. MAKING SIM.	.84
	INTERVIEW	.67
3RD LEVEL SUP.	1 DAY A.C.	.47
	TECHNICAL KNOWL.	.00

- 13 -

14

TABLE 12 (continued)

GMIB CORRELATIONS WITH OTHER SELECTION DEVICES

SUPERVISORY LEVEL	SELECTION PROCESS	CORRELATION
2ND LEVEL SUP.	LGD - ASSIGNED	.61
	LGD - UNASSIGNED	.61
	ORAL PRESENT.	- .47
4TH LEVEL SUP.	INTERVIEW	.31
1ST LEVEL SUP.	ANALYSIS/REPORT	.70
3RD LEVEL SUP.	2 UNASSIGNED LGD'S	.72
2ND LEVEL SUP.	2 UNASSIGNED LGD'S	.58
3RD LEVEL SUP.	2 UNASSIGNED LGD'S	.15
3RD LEVEL SUP. (2 DAY A.C.)	JUDGMENT	.90
	LEADERSHIP	.87
	ANALYSIS	.79
	DECISIVENESS	.86
	INTERPERSONAL	.71

Summary and Discussion

The GMIB is a new approach to in-basket testing. Items are scored individually based on explicit scoring guidance. The items are designed to test candidate skills on commonly occurring management situations and are not tied to any particular management position. The scoring guidance was developed based on the application of prevailing management theory and sound management principles to commonly occurring management situations or problems.

The GMIB can be scored in a highly reliable and efficient manner. Across four studies of inter-rater reliability, the lowest obtained reliability was .87. Three studies found inter-rater reliability to be higher than .90. This makes the item-by-item scoring approach extremely attractive in comparison to traditional approaches.

Significant and substantial validity coefficients were found for predicting supervisory ratings of job performance. The performance criteria that were included are routinely identified in job analysis studies as critical to success in management positions. Based on the overall performance measures derived from these ratings, the GMIB has true validity in the .40 to .44 range. The highest validity is for predicting performance in relation to "all" other employees in the sample. The lowest true validities were for predicting subjective overall ratings of performance. These validities were .35 and .37, for first and second level supervisors, respectively.

Total scores on the GMIB were significantly correlated with all six performance dimensions. The estimated true validities for the next-higher-level supervisors ("all" levels) had the highest range of coefficients, from .28 to .56. Across the six dimensions, Fisher's r to z transformation yields a mean true validity of .42.

The inclusion of a criterion measure in which employees are rated in relation to all other employees at the same level but regardless of job classification is unusual for validation studies -- as is obtaining ratings on employees in relation to others regardless of organizational level. Typically, validation studies are based on rating employees in relation to others at their same level and in their same classification. The fact that obtained validities using both the "same" level and "all" levels criterion measures were substantial and highly significant indicates that the General Management In-Basket is neither "position-bound" nor "level-bound" -- at least within the range of employees and organizational levels included in this study.

The results of the factor analysis make it possible to profile candidates on their particular strengths and weaknesses. Traditional in-basket scoring approaches rate candidates on dimensions and attempt to achieve reliable profile information, although assessors frequently experience difficulties in clearly distinguishing between dimensions. Due to the item-by-item scoring approach of the General Management In-Basket, mathematically independent factor (dimension) scores can be readily generated for each candidate. This approach avoids the problems inherent in the traditional approach of attempting to make clear distinctions among dimensions which are often highly related and therefore not readily susceptible to such distinctions.

In summary, there are a number of problems associated with the traditional approach to in-baskets. Raters have great difficulty in sorting out the wealth of behavioral information and making reliable ratings on assessment dimensions. Due to concerns for reliability, two or more raters are usually needed to score such in-baskets. Given the high inter-rater reliability of the GMIB, however, only a single rater is necessary. Moreover, the scoring time of the GMIB is minimal in comparison to traditional in-baskets.

To date, there have been no challenges or negative feedback with regard to the content of the General Management In-Basket. It has been used as a selection tool for both small and large organizations with no challenges to its relevance as a measure of success in management.

We should remember that candidates for management positions are typically a well-educated group. "Face validity", in the sense of placing candidates in a scenario equivalent to the target position, is not likely to be the deciding issue with such candidates. Much more important would seem to be the relevance and realism of the problems which are posed in the in-basket.

- 15 -

16

REFERENCES

Cronbach, L.J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Gaugler, B.B., Rosenthal, G.C., Thornton III, G.C., and Bentson, C. Meta-analysis of assessment center validity. Journal of Applied Psychology, 1987, 72, 493-511.

Gorsuch, Richard L. Factor Analysis. W. B. Saunders Company, Philadelphia, 1974.

Guilford, J. P. Psychometric Methods. McGraw-Hill, New York, 1954.
Hinrichs, J. R. and Haanpera, S. Reliability of measurement in situational exercises: an assessment of the assessment center method. Personnel Psychology, 1976, 29, 31-40.

Kesselman, G. A., Lopez, F. M., and Lopez, F. E. The development and validation of a self-report scored in-basket test in an assessment center setting. Public Personnel Management Journal, 1982, 11, 228-238.

Kim, J. Factor analysis. In Nie, N.J., Hull, C.H., Jenkins, J.G., Steinbrenner, K., Brent, D.H. (Eds.), SPSS: Statistical Package for the Social Sciences, pp. 468-514. McGraw-Hill, New York, 1975.

Kraus, J. C. The multiple-choice in-basket exercise as developed and used by the N. J. department of civil service. Paper presented at the 1986 International Personel Management Association Assessment Council annual conference, San Francisco, Ca.

Lopez, F. M. Evaluating executive decision-making: The in-basket technique. American Management Association, Inc. (AMA Research Study 75), 1966.

Schippmann, J. S., Prien, E. P., and Katz, J. A. Reliability and validity of in-basket performance measures. Personnel Psychology, 1990, 43, 837-859.

Thornton, G. C. III and Byham, W. C. Assessment Centers and Managerial Performance. Academic Press, New York, 1982.

Wollowick, H. B. and McNamara, W. J. Relationship of components of an assessment center to management success. Journal of Applied Psychology, 1969, 53, 348-352.