NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

A Study of a Technique for Validating Learning Hierarchies

by

Wai-Man Chan

Thesis presented to the School of Graduate Studies of the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Arts (Education)

Ottawa, Ontario, 1979

ACKNOWLEDGEMENT

In preparing this thesis, valuable advice and guidance were given to the researcher by Dr. Marvin Boss and Dr. Jean-Paul Dionne. Dr. Boss acted as thesis committee chairman.

The researcher is indebted to the Ontario Ministry of Colleges and Universities for its financial support in terms of an Ontario Graduate Scholarship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

One of the techniques for testing the validity of a learning hierarchy was carefully examined. This validating technique, called MAX, was developed by Owston. The maximum likelihood method for parameter estimation was used in the MAX technique.

With computer simulated data, the performance of the MAX technique was compared(by Owston) with that of other validating techniques. It was claimed that the performance of MAX was consistently better.

However, a number of weaknesses were identified in the original procedure for applying the MAX technique. These weaknesses are described near the end of the first chapter, after a brief review of the development of the MAX technique has been made.

Because of the weaknesses in the original procedure, modifications to this procedure appear necessary. Two modified procedures were developed and are discussed in Chapter II. These two procedures are basically the same as the original procedure, except that, the consequence of

ignoring the weaknesses is considered. The design for data collection is similar to that used in the original study. Computer simulated data are also used in this study. The results obtained from using the two modified procedures are presented and discussed in the last chapter. Some technical details involved in this study are shown in appendix A to E.

Chapter I

CRITICAL REVIEW OF THE LITERATURE

The concept of learning hierarchies has been used in different areas in the field of education. Among the techniques available to validate the logically derived learning hierarchies, Owston (1978) demonstrated that his MAX technique performed better than other techniques. However, there are some weaknesses in his application of the technique that may have biased the research findings.

In this chapter, after a brief review of the concepts relating to learning hierarchies and their validation methodologies, descriptions are made of the MAX technique and the general design of Owston's study. Then, weaknesses of the technique are discussed. In the last section of the chapter the research problem is discussed.

1.1 LEARNING HIERARCHIES AND THEIR VALIDATION

A learning hierarchy is a network of well-defined intellectual skills organized into levels, with the terminal skill(s) at the top level and subordinate skill(s) at the lower level(s); any subordinate skill is hypothesized to be

a prerequisite to the directly connected skill(s) in the higher level(s). The simplest (Durell, 1974, p. 4) learning hierarchy consists of two levels, with one skill in each level. A complicated hierarchy may be composed of numerous connected pairs (Owston, 1978, p. 20 and White, 1974, p. 61) of skills, each pair being equivalent to the simplest hierarchy.

The concept of learning hierarchies as networks of prerequisite relationships of intellectual skills was first introduced by Gagne (Gagne, 1962 and Gagne & White, 1974, p. 19). Uses of the concept are reported (Durell, 1974, p. 1 and Owston, 1978, p. 8-10) in the designing of individualized instruction, mastery learning sequences, computer-based instructional systems, and in subject areas (Gagne & White, 1974, p. 21) such as mathematics and science, where quantitative skills are essential.

To generate a learning hierarchy, a curriculum designer begins with the terminal skill(s) to be learned and asks the question (Gagne, 1962, p. 358) "What would the individual(s) have to be able to do in order that he(they) can attain successful performance on this task (skill expressed in behavioral terms)?". The answer to the question leads to the identification of immediately prerequisite skill(s). By reapplying the question to each of those newly identified

skills not yet possessed by the individual(s), subordinate skill(s) of lower level(s) are derived. The questioning process is repeated on the derived skills until (Cotton et, al., 1977, p. 190) the last identified skill can no longer be decomposed into simpler prerequisite skills which are not already possessed by the prospective learner(s).

For the simplest learning hierarchy, the process of its generation is as follows: On applying the question to the terminal skill, the curriculum designer can identify only one prerequisite skill the individual(s) has(have) to learn. On reapplying the question to the prerequisite skill, he finds that the prospective learner(s) does (do) not need to learn other subordinate skills for the attainment of the prerequisite skill. The process of questioning will stop, with two skills being in the learning hierarchy.

There is no certainty that hierarchies, derived by using the above procedures, are valid (Durell, 1974, p. 2). Two approaches are available to test the validity. The first approach is based upon the assumption that in training a learner on a certain skill, his training on an immediately subordinate skill will lead to positive transfer (Gagne 1965, p. 233 and Gagne, 1968) to the learning of the present skill. The second approach is based on the performance patterns of the learners. It involves utilizing the

performance pattern to test the hypothesis that, except by chance, learners cannot be successful in answering questions on a certain skill if they are not successful in answering questions on the prerequisite skill(s).

The first approach requires (White, 1973, p. 373) setting up a large number of experimental groups and instructional interventions even for a fairly simple hierarchy to be validated. For example, a hierarchy consisting of three skills may require setting up 3! (or 6) experimental groups; a hierarchy of k skills may require k! groups. Hence, this approach is not commonly used. The second approach is less rigorous than the first one in ensuring positive transfer of learned capacities to a skill from its immediate prerequisite skill(s). However, due to the practical importance of requiring less time and instructional intervention, the second approach is widely used (Durell, 1974, p. 2 and White, 1973, p. 373).

Two classes (Owston, 1978, p. 29) of techniques for validating learning hierarchies may be identified in the second approach. They are deterministic techniques and probabilistic techniques. Deterministic techniques (White, 1974, p. 65) have one or both of the following types of short-comings. These shortcomings include the ignoring of errors of measurements in analysing test results of the

learners, and judging the criterion of skill possession in terms of the number of test items answered correctly.

White & Clark devised a probabilistic technique (White, 1974, p. 64 and White & Clark, 1973) which is free from the above short-comings. This technique is useful for validating two-skill learning hierarchies (the White & Clark Model of learning hierarchies; to be discussed in the next section). However, the White & Clark technique has been criticized (Owston, 1978, p. 62) as being weak in the estimation of parameters contained in the White & Clark Model. Information available for parameter estimation has not been fully utilized in this technique.

Owston (1978, p. 62-79), using some of the assumptions of the White & Clark Model, developed a probabilistic validating technique, called the MAX technique. This technique involves using the maximum likelihood method for estimating population parameters.

No matter whether it is a deterministic or a probabilistic technique, there may always be decision errors (Owston, 1978, p. 29-30) involved in using the technique. For a hierarchy that is valid, the application of a technique may lead to an indication of invalidity. For an invalid hierarchy, the result from applying the technique may cause the researcher to think that the hierarchy is valid.

Owston considered that the accuracy (Owston, 1978, p. 107) of a technique may be judged by the extent of its simultaneously satisfying predetermined probabilities for each type of decision error. The emphasis on simultaneity may be explained with an extreme example; a technique that indicates correctly all the hierarchies that are valid and indicates incorrectly all the hierarchies that are invalid would be useless as would one that performs in an opposite manner.

In trying to establish the accuracy of a technique or to compare two or more different techniques, one more complication exists; in real life, one never knows whether a derived hierarchy is valid or invalid, for if the underlying validity were known, no testing would be required (Durell, 1974, p. 3). Computer simulation provides a means of generating simulated learner performance patterns related to hierarchies of "known" validity. With computer simulation, the researcher can determine the validity of hierarchies by specifying population parameters. Also, he can easily vary the size of samples, size of measurement errors, difficulty levels, validity of the hierarchy, etc..

Using computer simulated data, Owston (1978) studied the influences of measurement errors, skill difficulty levels and sample sizes on the accuracies of the MAX

technique, the White & Clark technique, and two (Owston, 1978, p. 81-82) deterministic techniques, the Walbesser Ratios and the Difference Ratio. The learning hierarchies Owston generated were the simplest (Owston, 1978, p. 20 and White, 1974, p. 65) ones with only two skills involved. Based on his findings, he concluded (Owston, 1978, p. 141) that the MAX technique was more accurate than the other techniques over different skill difficulty levels, measurement errors and sample sizes commonly encountered in research.

To have a better understanding of the MAX technique, it is important to know the White & Clark Model upon which the MAX technique is based. It is also important to have an understanding of Owston's experimental design used in testing the accuracy of MAX. In the next few sections, the White & Clark Model and the MAX technique will be described, followed by a critical review of Owston's study and his technique.

1.2 THE WHITE & CLARK MODEL

White & Clark developed a probabilistic model (White & Clark, 1973 and Owston, 1978, p. 53-62) to test the validity of a hierarchical connection between a pair of skills, called skill I and skill II, with skill I being the subordinate skill. Two test items were used to test each skill. Measurement errors were considered.

In a population of students four mutually exclusive groups can be formed. The four groups, respectively, contain students who possess, neither skill I nor skill II, only skill I, only skill II, both skill I & skill II. The corresponding proportions of students within the population are denoted by P_O, P_I, P_{II} and P_B.

For skill I, the two test items are assumed to be very similar to, but independent of, each other. This means that, for any student, the chance of correctly answering the first item is the same as the chance of correctly answering the second item; however, the actual outcomes (correct or incorrect) of answering the two items, are independent of each other. This is also true for the two items measuring skill II.

Based on the above assumption, probabilities of correctly answering the test items may be associated with each student in accordance with his state of skill possession. For a student who possesses skill I (he is either from the P_I or P_B group), his probability, of correctly answering each of the two independent skill I test items, is indicated by θ_a. For a student who does not possess skill I (he is either from the P_O or P_{II} group), the corresponding probability of correctly answering each of the two skill I test items is indicated by θ_b. Similarly, two

probabilities are related to skill II test items. For a student who possesses skill II (he is either from the P_{II} or P_B group), his probability, of correctly answering each of the two independent skill II test items, is denoted by θ_c. For a student who does not possess skill II (he is either from the P_O or P_I group), the corresponding probability of correctly answering the two items is denoted by θ_d.

Note that θ_b and θ_d are related to the measurement errors of correct guessing on skill I and skill II test items respectively. Furthermore, note that $(1-\theta_a)$ and $(1-\theta_c)$ are related to the measurement errors of blundering(failing to answering correctly when possessing the skill), respectively, on skill I and skill II test items.

With the four population proportions, P_O, P_I, P_{II} and P_B, and the four probabilities of obtaining a correct answer, θ_a, θ_b, θ_c and θ_d, the White & Clark Model can be formed. Table I shows the model which contains nine cells and the corresponding theoretical probability functions associated with each cell.

Corresponding to the theoretical model shown in Table I, a 3x3 table of observed response patterns can be obtained from a sample of students who respond to the skill I and skill II test items.

TABLE I

Probabilities of Members of the Sample being Classified in Particular Cells *

Skill II

Skill I: Questions correct	0	1	2	
2	cell 1 $P_{20} = p_1$ $= P_0\theta_b^2(1-\theta_d)^2$ $+P_I\theta_a^2(1-\theta_d)^2$ $+P_{II}\theta_b^2(1-\theta_c)^2$ $+P_B\theta_a^2(1-\theta_c)^2$	cell 2 $P_{21} = p_2$ $=2P_0\theta_b^2\theta_d(1-\theta_d)$ $+2P_I\theta_a^2\theta_d(1-\theta_d)$ $+2P_{II}\theta_b^2\theta_c(1-\theta_c)$ $+2P_B\theta_a^2\theta_c(1-\theta_c)$	cell 3 $P_{22} = p_3$ $=P_0\theta_b^2\theta_d^2$ $+P_I\theta_a^2\theta_d^2$ $+P_{II}\theta_b^2\theta_c^2$ $+P_B\theta_a^2\theta_c^2$	a
1	cell 4 $P_{10} = p_4$ $=2P_0\theta_b(1-\theta_b)(1-\theta_d)^2$ $+2P_I\theta_a(1-\theta_a)(1-\theta_d)^2$ $+2P_{II}\theta_b(1-\theta_b)(1-\theta_c)^2$ $+2P_B\theta_a(1-\theta_a)(1-\theta_c)^2$	cell 5 $P_{11} = p_5$ $=4P_0\theta_b(1-\theta_b)\theta_d(1-\theta_d)$ $+4P_I\theta_a(1-\theta_a)\theta_d(1-\theta_d)$ $+4P_{II}\theta_b(1-\theta_b)\theta_c(1-\theta_c)$ $+4P_B\theta_a(1-\theta_a)\theta_c(1-\theta_c)$	cell 6 $P_{12} = p_6$ $=2P_0\theta_b(1-\theta_b)\theta_d^2$ $+2P_I\theta_a(1-\theta_a)\theta_d^2$ $+2P_{II}\theta_b(1-\theta_b)\theta_c^2$ $+2P_B\theta_a(1-\theta_a)\theta_c^2$	b
0	cell 7 $P_{00} = p_7$ $=P_0(1-\theta_b)^2(1-\theta_d)^2$ $+P_I(1-\theta_a)^2(1-\theta_d)^2$ $+P_{II}(1-\theta_b)^2(1-\theta_c)^2$ $+P_B(1-\theta_a)^2(1-\theta_c)^2$	cell 8 $P_{01} = p_8$ $=2P_0(1-\theta_b)^2\theta_d(1-\theta_d)$ $+2P_I(1-\theta_a)^2\theta_d(1-\theta_d)$ $+2P_{II}(1-\theta_b)^2\theta_c(1-\theta_c)$ $+2P_B(1-\theta_a)^2\theta_c(1-\theta_c)$	cell 9 $P_{02} = p_9$ $=P_0(1-\theta_b)^2\theta_d^2$ $+P_I(1-\theta_a)^2\theta_d^2$ $+P_{II}(1-\theta_b)^2\theta_c^2$ $+P_B(1-\theta_a)^2\theta_c^2$	c
	f	e	d	

* Reproduced from White & Clark (1973, p. 77);

a, b, c, d, e, f are marginal frequencies.

Using the theoretical model and the observed 3x3 table together, it is possible to derive useful relationships for inferring validity of the learning hierarchy characterized by the population.

The MAX technique, to be described next, is one of the validating techniques based on the White & Clark Model of two-skills learning hierarchy.

1.3 THE MAX TECHNIQUE

The MAX technique (Owston, 1978, p. 62-79) consists in producing an interval estimate of the parameter P_{II}, which is the proportion of the population that possess the superordinate skill without possessing the subordinate skill. For a valid hierarchy P_{II} is equal to zero; for an invalid hierarchy P_{II} is larger than zero(but less than 1). The estimation is obtained through the application of the method of maximum likelihood. If the confidence interval includes zero, it is concluded that the hierarchy is valid, otherwise it is concluded that the hierarchy is invalid.

In the White & Clark Model, of the eight parameters (P_O, P_I, P_{II}, P_B, θ_a, θ_b, θ_c, θ_d), only seven have to be estimated because P_B can be expressed as $1-(P_O+P_I+P_{II})$. For convenience in deriving formulae to show the estimation process, new symbols are used such that $P_O = \theta_1$, $P_I = \theta_2$, $P_{II} = \theta_3$, $\theta_a = \theta_4$, $\theta_b = \theta_5$, $\theta_c = \theta_6$, $\theta_d = \theta_7$.

It is assumed in Rao's approach, as presented by Owston, that in a random sample of n subjects, the observed frequencies in the nine cells in Table I happen to be (f_1, f_2, ..., f_8, f_9). The probability of occurrence of such an observed pattern can be expressed as a multinomial distribution function called the likelihood function (L). Symbolically,

$$L = \frac{n!}{\prod_{i=1}^{9} f_i!} \prod_{i=1}^{9} p_i^{f_i} \qquad (1)$$

where $0 \leq f_i \leq n, \sum_{i=1}^{9} f_i = n$, and p_i is a function of the seven parameters, θ_1 to θ_7.

The seven parameter estimates that maximize L are called the maximum likelihood estimates (m.l.e.) of the parameters.

To simplify the derivation, the natural logarithm Log L is used; since it is a monotonic (Owston, 1978, p. 66 and Allendoerfer & Oakley, 1963, p. 209) increasing function of L, the estimates that maximize L will also maximize Log L. To find the m.l.e. from Log L, a scoring method can be used. Corresponding to the seven parameters to be estimated, there are seven efficient scores, the j^{th} efficient score (Rao, 1968, p. 302-305) is defined as the partial derivative of Log L with respect to θ_j, that is

$$S_j = \frac{\partial \text{Log } L}{\partial \theta_j} \qquad (2)$$

where $$\frac{\partial \text{Log } L}{\partial \theta_j} = \sum_{i=1}^{9} \frac{f_i \partial p_i}{p_i \partial \theta_j} \qquad j = 1, 2, \ldots, 7 \qquad (3)$$

Theoretically, by setting the seven efficient scores in equation (2) to zeros, and solving for θ_1 to θ_7, the solution (Rao, 1968, p. 302) thus obtained will be the maximum likelihood estimator. In practice, for estimates that are functions of several parameters, it is too complicated to proceed with that approach.

By using the scoring procedures, an initial set (Rao, 1968, p. 302) of parameter estimates $\hat{\theta}^o = (\hat{\theta}_1^o, \hat{\theta}_2^o, \ldots, \hat{\theta}_7^o)$ is used as a first approximation of the solution. The efficient scores in equation (2) are expanded about the initial set via the Taylor series. Only the first order terms in $\overset{o}{\delta\theta}_j$ are retained, where $\overset{o}{\delta\theta}_j = \theta_j - \hat{\theta}_j^o$ is the deviation of the j^{th} initial parameter estimate from the j^{th} maximum likelihood estimate($\overset{o}{\delta\theta}_j$ can be interpreted as a correction of the initial approximation). Following Taylor's expansion, the efficient scores are equated to zero, the following equation is obtained:

$$0 = S_j\Big|_a + \overset{o}{\delta\theta}_1 \frac{\partial^2 \text{Log } L}{\partial\theta_1 \partial\theta_j}\Big|_a + \cdots + \overset{o}{\delta\theta}_6 \frac{\partial^2 \text{Log } L}{\partial\theta_6 \partial\theta_j}\Big|_a + \overset{o}{\delta\theta}_7 \frac{\partial^2 \text{Log } L}{\partial\theta_7 \partial\theta_j}\Big|_a + \cdots \qquad (4)$$

where j=1,7 and the symbol $\Big|_a$ implies evaluating the immediately preceding function at the point "a"; in this case "a" is the point $(\hat{\theta}_1^o, \hat{\theta}_2^o, \ldots, \hat{\theta}_7^o)$.

For large samples (Rao, 1968, p. 302, and Owston, 1978, p. 68), the information for score j, I_{jk}, can be defined as

$$-\frac{\partial^2 \text{Log } L}{\partial\theta_j \partial\theta_k} \quad \text{where} \quad \begin{matrix} k = 1,2,\ldots,7 \\ j = 1,2,\ldots,7 \end{matrix} \tag{5}$$

that is $$I_{jk} = -\frac{\partial^2 \text{Log } L}{\partial\theta_j \partial\theta_k} = n \sum_{i=1}^{9} \frac{1}{p_i} \frac{\partial p_i}{\partial\theta_j} \frac{\partial p_i}{\partial\theta_k} \tag{6}$$

Using a small circle on the top of the symbols to indicate initial values, equation (4) can be rewritten as

$$0 = \overset{o}{S}_j - \overset{o}{I}_{jk} \times \overset{o}{\delta\theta}_k + \cdots \tag{7}$$

where j,k=1,2,3,...,7. By rearranging the terms in (7),

$$\overset{o}{\delta\theta}_k = \overset{o\,-1}{I}_{jk} \times \overset{o}{S}_{jk} \tag{8}$$

In matrix form, equation (8) becomes

$$\begin{bmatrix} \theta_1-\hat{\theta}_1 \\ \theta_2-\hat{\theta}_2 \\ \vdots \\ \theta_7-\hat{\theta}_7 \end{bmatrix} = \begin{bmatrix} \overset{o}{\delta\theta}_1 \\ \overset{o}{\delta\theta}_2 \\ \vdots \\ \overset{o}{\delta\theta}_7 \end{bmatrix} = \begin{bmatrix} \overset{o}{I}_{11} & \overset{o}{I}_{12} & \cdots & \overset{o}{I}_{17} \\ \overset{o}{I}_{21} & \overset{o}{I}_{22} & \cdots & \overset{o}{I}_{27} \\ \vdots & \vdots & \cdots & \vdots \\ \overset{o}{I}_{71} & \overset{o}{I}_{72} & \cdots & \overset{o}{I}_{77} \end{bmatrix}^{-1} \begin{bmatrix} \overset{o}{S}_1 \\ \overset{o}{S}_2 \\ \vdots \\ \overset{o}{S}_7 \end{bmatrix} \tag{9}$$

The left hand side of equation (9) represents a column vector of additive corrections to the initial set of parameter estimates. This column vector is found by calculating the elements of the 7x7 information matrix using equation (6), evaluated at initial parameter estimate values, then obtaining the inverse of the 7x7 matrix and

multiplying the inverted matrix by a column vector of efficient scores calculated from equation (2).

With the initial set of parameter estimates and the derived vector of additive corrections, a new set of parameter estimates closer to the m.l.e. can be found by adding the initial set and the vector of corrections. Let the new set of estimates be $\hat{\theta}^1$, then from (9)

$$\begin{bmatrix} \hat{\theta}_1^1 \\ \hat{\theta}_2^1 \\ \vdots \\ \hat{\theta}_7^1 \end{bmatrix} = \begin{bmatrix} \hat{\theta}_1^o \\ \hat{\theta}_2^o \\ \vdots \\ \hat{\theta}_7^o \end{bmatrix} + \begin{bmatrix} {}^o\delta\theta_1 \\ {}^o\delta\theta_2 \\ \vdots \\ {}^o\delta\theta_7 \end{bmatrix} \qquad (10)$$

Repeating the process with the new approximations, i.e. substituting $\hat{\theta}_j^1$ values into equation (3), (6) and (9), a new set of additive corrections can be obtained. The process is repeated (Rao, 1968, p. 305 and Owston, 1978, p. 70) until the newly derived additive corrections are sufficiently small, for example, less than 0.0001. The final set of parameter estimates, with stable values, are good approximations of the maximum likelihood estimates.

After the m.l.e. are found, their variances (Elandt-Johnson, 1971, p. 306) can also be obtained. The variance of the j^{th} parameter estimate, $\text{var}\hat{\theta}_j$, is approximately equal to the value of the j^{th} diagonal element in the inverted information matrix with elements calculated at values of the

maximum likelihood estimates. Since the parameter estimate of interest in the MAX technique is $\hat{P}_{II} = \hat{\theta}_3$, its variance, $\text{var}\hat{\theta}_3$, is approximated by the (3,3)th element of the 7x7 inverted information matrix just mentioned.

The MAX technique is based on the fact that maximum likelihood estimates are approximately normally distributed (Owston, 1978, p. 78). A confidence interval can be built around $\hat{\theta}_3$ at a level of confidence of $(1-\alpha)$, such that,

$$\hat{\theta}_3 - Z_{\alpha/2}\sqrt{\text{var}\hat{\theta}_3} < \theta_3 < \hat{\theta}_3 + Z_{\alpha/2}\sqrt{\text{var}\hat{\theta}_3} \qquad (11)$$

$$\text{or} \quad \hat{P}_{II} - Z_{\alpha/2}\sqrt{\text{var}\hat{P}_{II}} < P_{II} < \hat{P}_{II} + Z_{\alpha/2}\sqrt{\text{var}\hat{P}_{II}} \qquad (12)$$

where $Z_{\alpha/2}$ is the standard normal deviate (Miller & Freund, 1965, p. 146-148) with the fraction of the area under the normal curve and to the right being equal to $\alpha/2$.

To illustrate the process of building a confidence interval around $\hat{P}_{II}$ and the subsequent decision on the validity of the hierarchy, part of the example given in Owston's thesis is used here. In Table II, the initial set of parameters is identified by the iteration number 0. This initial set is obtained from the 3x3 table of observed frequencies by using the marginal frequency method (devised by White & Clark; see Appendix B). The transition from one iteration to another is attained by the method of scoring.

Table II

Parameter Estimation Using the Method of Scoring *

Iteration Number	$\hat{P}_O$	$\hat{P}_I$	$\hat{P}_{II}$	$\hat{P}_B$	$\hat{\theta}_a$	$\hat{\theta}_b$	$\hat{\theta}_c$	$\hat{\theta}_d$
0	.0886	.3318	.0000	.5796	.9731	.0000	1.0000	.1929
1	.0959	.2351	.0055	.6635	.9799	.0694	.9389	.1085
2	.1006	.2351	.0068	.6584	.9829	.0910	.9432	.0988
3	.0996	.2308	.0070	.6624	.9824	.0866	.9406	.0938
4	.1001	.2316	.0070	.6611	.9827	.0886	.9414	.0954
5	.0996	.2312	.0070	.6617	.9826	.0878	.9410	.0947
6	.1000	.2314	.0070	.6615	.9827	.0881	.9412	.0949
7	.1000	.2314	.0070	.6616	.9827	.0880	.9411	.0949
8	.1000	.2314	.0070	.6615	.9827	.0881	.9412	.0949
9	.1000	.2314	.0070	.6615	.9827	.0881	.9412	.0949
10	.1000	.2314	.0070	.6615	.9827	.0881	.9412	.0949

* Reproduced from Owston, p. 77.

After the 8^{th} iteration, the new additive corrections are so small (less than .0001 here), that the 8^{th} set can be assumed to be the set of maximum likelihood estimates. $\hat{P}_{II}$ has the value of .0070, and from the inverse of the information matrix calculated from the m.l.e. (maximum liklihood estimates), the estimated variance of $\hat{P}_{II}$ has a value of 7.451×10^{-5} .

For a 95% confidence interval around $\hat{P}_{II}$, $Z_{\alpha/2}$ is about 1.96 . Substituting $Z_{\alpha/2}$, $\hat{P}_{II}$ and var $\hat{P}_{II}$ values into equation (12), the following interval is obtained,

$$-0.00944 < P_{II} < 0.02394 .$$

That is, the population value of P_{II} is estimated to lie between (-0.00944) and (0.02394). Since zero is included in this interval, it is concluded that, the hierarchy is valid, with skill I being a prerequisite to skill II.

1.4 OWSTON'S DESIGN AND FINDINGS ON MAX

The theoretical basis of the MAX technique has been discussed. An example was presented to illustrate how the MAX technique functions.

In this section, Owston's design to test the MAX technique is described. Also, his findings are discussed.

1.4.1 Generating random samples

In real life one never knows whether a learning hierarchy is valid or invalid. This is a complication in testing the performance of a validating technique. In view of this, the testing problem can be approached via computer simulation.

Figure 1 shows the computer flowchart, used by Owston, for random generation of simulated subjects' responses to the test items related to the two-skill hierarchy described in the White & Clark Model.

The most important symbols, used in the flowchart, are those related to the White & Clark Model. Recall that, according to this model, four proportions exist in a population of subjects. The four proportions, P_O, P_I, P_{II} and P_B, correspond to four groups of subjects who possess "neither skill I nor skill II", only skill I, only skill II, and both skill I and skill II, respectively. Also recall that in this model 2 test items (or questions) per skill are used to test skill possession. Measurement error of correct guessing on skill I and skill II test items are indicated by θ_b and θ_d respectively. The measurement errors of blundering on skill I and skill II test items are denoted by $(1-\theta_a)$ and $(1-\theta_c)$ respectively. Note that θ_a is the probability of correctly answering skill I test items, for

Figure 1

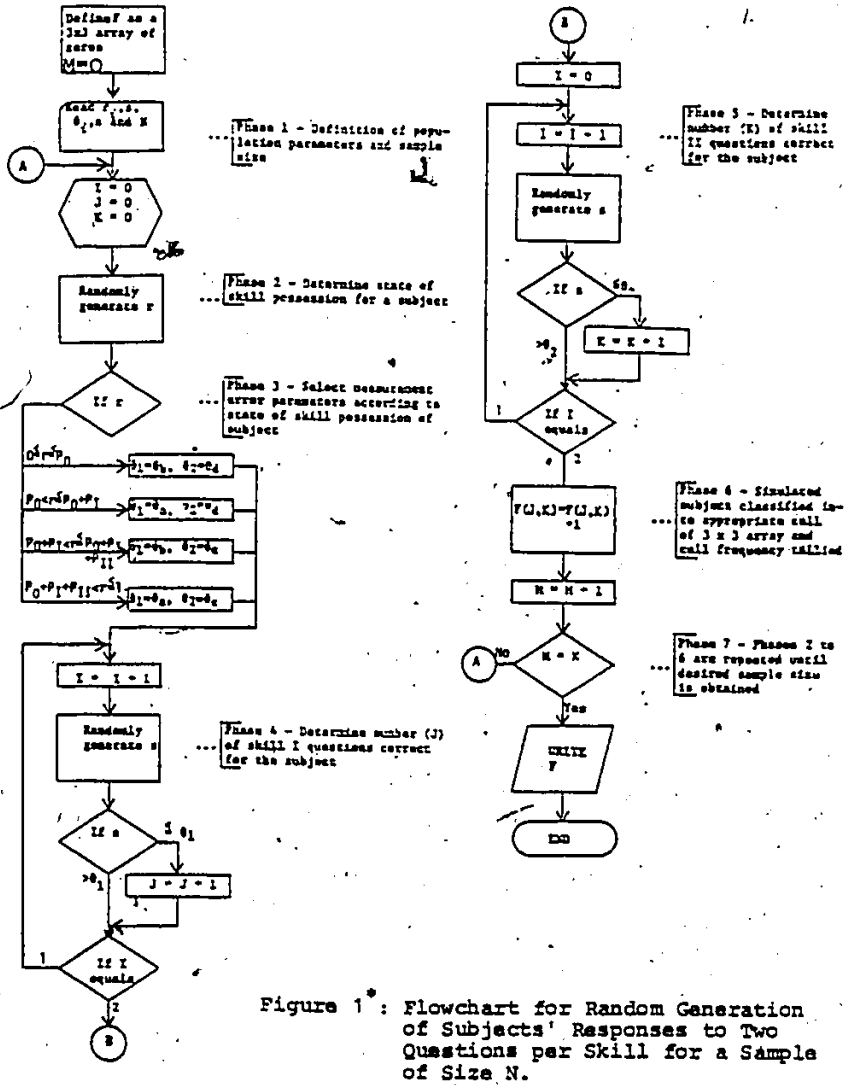


Figure 1*: Flowchart for Random Generation of Subjects' Responses to Two Questions per Skill for a Sample of Size N.

* Reproduced from Owston, p. 87.

those subjects possessing skill I. θ_c is the probability of correctly answering skill II test items, for those subjects possessing skill II.

Notice that there are seven phases indicated in the flowchart. Beginning with phase 1, the four P's and four θ's, just described, are assigned numerical values. Owston specified different sets of P's to represent valid and invalid hierarchies at different levels of skill difficulty for skill I and skill II. He also specified different sets of θ's to represent various levels of measurement errors. (The details of specification of P's and θ's will be discussed later in this section). Now, assume that four P and four θ values are specified to describe a population of "subjects" related to a hierarchy of certain level of skill difficulty and of a certain level of measurement errors. Then, a simulated sample of size N can be generated via the algorithm schematically represented in phase 2 to 7 of the flowchart.

Basically, it is through computer generation of a random number that Owston simulated a "subject" and assigned the "subject" to one of the four groups described in the White & Clark Model. Then, according to which group the "subject" was being assigned, the probabilities of correctly answering skill I and skill II test items, respectively were

determined. Futhermore, the "actual" responses of the "subject" to the test items were also simulated via a random number generation algorithm represented in phases 4 and 5.

The simulated response pattern of the first "subject" was then recorded in a 3x3 table. The process, in phase 2 to phase 6, was repeated N-1 times. By the end of this iterative process, the 3x3 table would contain the response patterns, of the N "subjects" in the sample, to the two test items per skill.

Then, Owston applied the MAX technique to the 3x3 pattern of simulated responses. The result from applying MAX would be a validity indication. Whether the indication is correct or incorrect depends on the validity of the hierarchy characterized by the "population of subjects".

1.4.2 Estimating risk of decision errors

Actually, using a set of population parameters (P's and θ's) a large number of samples, of differing sizes, can be generated. Owston selected four levels of sample sizes to be 25, 40, 100 and 400. At each level of sample size, 1000 samples were generated. The MAX technique was applied, to make validity decisions, on each sample generated. Two levels of decision criteria, the 95% confidence interval and the 90% confidence interval were used. The correct

decisions, at each decision criterion, were tallied. So, for a given set of population parameters, at certain level of sample size, the results, of generating 1000 samples and applying the MAX technique 1000 times, would be two tallied totals. The two tallied totals correspond to the number of correct decisions made at 95% confidence interval and 90% confidence interval, respectively, on the 1000 samples generated. Owston determined the total number of incorrect decisions, at each decision criterion, simply by subtracting each tallied total from 1000.

Recall that there are two types of decision errors that can be made. One type is related to incorrectly indicating valid hierarchies as "invalid", the "I-V" type. Another type is related to indicating invalid hierarchies as "valid", the "V-I" type. Owston estimated the risk of committing decision errors by expressing the total number of incorrect decisions as a fraction of 1000.

Hence, for a valid hierarchy, the risk of committing decision errors (due to the application of MAX) would be

$$P(I\text{-}V) = \frac{\text{total no. of incorrect decisions when hierarchy is valid}}{1000}$$

For an invalid hierarchy, the corresponding risk of committing decision errors would be

$$P(V\text{-}I) = \frac{\text{total no. of incorrect decision when hierarchy is invalid}}{1000}$$

1.4.3 Specifying population parameters

So far, in this section, discussions have been centered around one set of population parameters that characterizes a learning hierarchy with a certain level of skill difficulty and with a certain level of measurement errors. From this set of parameters, 1000 samples at certain sample sizes were hypothetically generated. A procedure for estimating the risk of committing decision errors was also outlined. However, the details of specification of different sets of parameters were not discussed. In the next few paragraphs, Owston's procedure for specifying 48 sets of population parameters is presented.

In Table III a schematical representation of Owston's design for testing the MAX technique is shown. Notice that the 48 sets of population parameter values are formed by combining the different levels of validity, skill difficulty, and measurement errors.

Table III

Schematical Representation of Owston's design for testing the MAI technique, over different decision criteria, sample sizes, measurement errors, skill difficulty levels and validity levels.

48 sets of population parameter values were formed by the following combinations:			Based on each set of parameter values, random samples were generated. 1000 samples were generated at each of the following 4 levels of sample size:	For each sample generated, MAI technique was applied to construct a confidence interval (C.I.) around $\hat{P}_{II}$, then it was applied make a validity decision. 2 levels of decision criteria were used for each sample. The 2 levels were:
3 levels of validity	4 levels of skill difficulty (skill I, skill II) or (D_I, D_{II})	4 levels of measurement errors of guessing (θ_b, θ_d)		
Valid	(0.95, 0.90)	0.05	25	95% C.I.
	(0.70, 0.50)	0.10	50	
Invalid (ϕ=0.00)				
	(0.40, 0.25)	0.20	100	
Invalid (ϕ=0.25)	(0.10, 0.05)	0.25	800	90% C.I.

First, Owston determined four levels of skill difficulty D_I for skill I as 0.95, 0.70, 0.40 and 0.10 . He argued that the performance of a validating technique might be related to the skill difficulty levels.

Then, he used his definition of a valid hierarchy to find the corresponding skill difficulty levels D_{II} for skill II as 0.90, 0.50, 0.25 and 0.05 . By Owston's definition of a valid hierarchy, skill II is more difficult than skill I, $P_{II}=0$, and the two skills are correlated with a phi correlation coefficient, ϕ, of approximately 0.68 . With this definition, once D_I is determined, D_{II} and the four P values can also be determined. (Apendix A shows the calculations involved in finding P values).

Corresponding to the four levels of skill difficulty pairs, (D_I, D_{II}), four sets of P's were determined. These four sets of P's were related to valid learning hierarchies.

Owston determined four other sets of P's to characterize invalid hierarchies with skill I and skill II being independent. He did this by using the same four levels of skill difficulties, (D_I, D_{II}), already determined, and by setting $\phi=0$.

Using the same four levels of skill difficulty, he also specified four sets of P's to characterize invalid

hierarchies with skill I and skill II being moderately correlated, $\phi = 0.25$.

After specifying 12 sets of P values, Owston determined four sets of measurement error probability values. For each of these four sets, he assumed that the probabilities of measurement error of blundering were constant and equal to 0.05 . That is, $(1-\theta_a)=(1-\theta_c)=0.05$; so, $\theta_a=0.95$ and $\theta_c=0.95$. He also assumed that the probabilities of guessing on skill I and skill II test items were equal. That is $\theta_b = \theta_d$. However, he allowed the probabilites of guessing to vary at four levels, with each level corresponding to one set of θ values. The four levels of guessing errors were 0.05, 0.10, 0.20 and 0.25 . In his discussions, .25 and .20 levels were related to the guessing probabilities of multiple choice types of items with four and five alternatives respectively. The two levels, .10 and .05, were related to guessing probabilities of short-answer types of items.

By combining 12 sets of P's with four sets of θ's, Owston obtained 48 sets of population parameters for random generation of simulated subjects' responses.

1.4.4 Generating datasets to examine the risks of decision errors

Recall that four levels of sample sizes were used for

each set of population parameters. At each level of sample size, 1000 random samples were generated. The MAX technique was applied to make 1000 validity decisions with 95% confidence intervals and simultaneously 1000 validity decisions with 90% confidence intervals.

Also recall the outlined procedure for estimating the risks, P(I-V) and P(V-I), of committing decision errors.

By combining 4 levels of sample sizes with the 48 sets of population parameters, 192 datasets were obtained for estimating the risks of decision errors under different conditions. Of these 192 datasets, 64 datasets were used to estimate P(I-V), 64 to estimate P(V-I) when $\phi=0$, another 64 to estimate P(V-I) when $\phi=0.25$.

1.4.5 Setting criteria on risks of committing decision errors

Of the two types of risks, P(I-V) and P(V-I), Owston considered P(V-I) as more serious because it was related to indicating an invalid hierarchy as "valid". So, he set up a more stringent criterion on P(V-I). Over the 128 datasets involving invalid hierarchies, Owston set a criterion of $P(V-I) \leq 0.05$; over the 64 datasets involving valid hierarchies, he set a criterion of $P(I-V) \leq 0.20$. Then, using 95% and 90% confidence intervals, Owston determined how often these criteria were met simultaneously on

populations that differed only in whether the hierarchy was valid or invalid.

In the next subsection, Owston's findings on the risks of decision errors, over the various pairs of datasets, will be presented.

4.4.6 Owston's findings on the risks of committing decision errors

Owston found that the risks of committing decision errors were related to the validity of hierarchy, decision criteria, sample sizes, probabilities of measurement errors, and the skill difficulty levels.

Recall that 95% and 90% confidence intervals were used as decision criteria. P(I-V) risks were found to be lower with 95% confidence interval than with 90% confidence interval. However, P(V-I) risks were higher with 95% confidence interval than with 90% confidence interval.

To compare the relative usefulness of 95% and 90% confidence intervals, Owston utilized the results obtained on the 64 datasets related to valid hierarchies of different levels of skill difficulty, and he also utilized the corresponding results on the 64 datasets related to invalid hierarchies with ϕ=0.25 (He did not use the results on the 64 datasets related to ϕ=0.00, because he argued that

datasets with ϕ=0.25 would provide more stringent comparisons).

With valid hierarchies, the P(I-V) values for 95% and 90% confidence intervals were less than the 0.20 criterion over all the 64 datasets. However, with invalid hierarchies(ϕ=0.25) the P(V-I) criterion of less than 0.05 was reached in 12 datasets using 95% confidence interval, and in 14 datasets using 90% confidence interval. In other words, the 90% confidence interval was found to be relatively more useful than the 90% confidence interval. Based on this finding, Owston suggested the use of shorter confidence intervals in future research.

Besides validity and decision criteria, sample sizes also played an important role in affecting the size of the error risks. In general, with all other things being equal, an increase in sample size corresponded to a decrease in the risk of committing decision errors.

Also, fewer decision errors were made with a decrease in the probability of measurement errors.

The only other variable is skill difficulty. With fixed levels on measurement errors, sample size and validity there was a general pattern on the size of risks over the four levels of skill difficulty. For both 95% and 90%

confidence intervals, the risks of P(V-I) and P(I-V) were generally higher for the two extreme difficulty levels of (.95, .90) and (.10, .05) than for the two intermediate levels of (.70, .50) and (.40, .25). Owston considered that this pattern might be due to the higher differences of $D_I - D_{II}$ at the two intermediate levels. For example, with the two intermediate levels, $D_I - D_{II}$ were .20 and .15 respectively. However, at the two extreme levels, the $D_I - D_{II}$ values were both .05 .

1.5 WEAKNESSES OF THE MAX TECHNIQUE

Owston's study of the MAX technique appears to be theoretically and practically logical. However, in a pilot study conducted to test the MAX technique, using Owston's computer program, various types of errors were identified. All these types of errors occurred when using the subroutine program in which Owston implemented the MAX technique.

In Table IV, 11 types of identified errors are shown. These error types were grouped into three error categories, I, II, and III. Category I error types, if not detected and by-passed, would cause computer interruptions to occur. It is assumed that Owston had somehow taken care of these errors. Error types in category II and category III would not cause any interruption to the running of the computer

Table IV: Error categories and error types identified when the MAX technique was pilot tested using Overton's computer program

Error Category	Category Description	Error type No.	Error type Label	Error Type Description
I	Error types in this category cause computer program interruptions. It is essential to check these types of errors.	1	"f=0"	Marginal frequency f=0. Leads to "division by zero".
		2	"p=0"	probability values P1 , P2 , ... P9, shown in Table I , approach zero.
		3	"Extreme"	Intermediate estimates take on extreme values.
		4	"SQRT5"	At the end of 5 iterations, variance of P_{II} may be a negative value. This cause computation of square root of a negative value.
		5	"SQRT2"	It is the same as "-SQRT5" except that convergence is checked after each iteration instead of at the end of the 5th iteration.
II	Error types in this category are related to non-convergence of parameter estimates.	1	"NCONV5"	"Non convergence" of estimates is detected by an "inadequate"* check at the end of 5 iterations.
		2	"PNCONV5"	"Inadequate" check indicates "Convergence" but proper** check indicates non-convergence.
		3	"NCONV10"	Non-convergence of estimates at the end of 10 iterations.
		4	"INTVAL"	Intermediate estimates take on unreasonable values.
III	This category of errors are associated with improperly convergent estimates	1	"-P_{II}"	Convergent P_{II} value happens to be a large negative value.
		2	"FINAL"	The final convergent estimates, other than P_{II}, are outside the normal (0,1) range.

* "inadequate" check for convergence is $\overrightarrow{\delta x} < .001$, where $\overrightarrow{\delta x}$ is the vector of corrections.

** "proper" check is $|\overrightarrow{\delta x}| < .001$

program. It is very likely that these error types had not received Owston's attention because in his computer program, no error checks were available to monitor these errors.

In error category I, the error type "f=0" would occur if a randomly generated response pattern had a form similar to that shown in Figure 2, where the marginal frequency f is equal to zero. The reason for the occurrence of this error type in the MAX technique is that Owston used the marginal frequency approach devised by White & Clark (see Appendix B for details) to obtain an initial set of parameter estimates. With this approach the value of f is used as denominator in some computational formulae. When f is equal to zero, computer interruption would take place. The identification of this error type leads the researcher to consider that one weakness of the MAX technique is that it may not always be possible to obtain a feasibe set of initial parameter estimates from the 3x3 patterns of observed responses.

This error type is expected to occur more often over datasets associated with small sample sizes and easy skills. Under the above conditions, the cells corresponding to either zero skill I or zero skill II questions correctly answered would more likely be empty. Hence it is also more likely that the marginal frequency value f is equal to zero.

No. of skill II questions correctly answered

No. of skill I questions correctly answered	0	1	2	
2	0	3	15	a
1	0	3	2	b
0	0	1	1	c
	f	e	d	

Figure 2: A 3x3 pattern in which the error type "f=0" would occur

The error type "p=0" occurred when the probability function(s) (such as $p_1, p_2, \ldots, p_9$), shown in Table I, had value(s) very close to zero. Computer interruption may take place due to "memory overflow" as a result of dividing some expressions by these function values. This error type is expected to occur more often in datasets with low measurement errors. Assume that the measurement errors were so low that there were no measurement errors at all (i.e. $\theta_b=\theta_d=0$, and $\theta_a=\theta_c=1$); then, in the White & Clark Model (shown in Table I) only the four corner cells (cells 1, 3, 7, 9) may contain non-zero probability functions. The other five probability functions (p_2, p_4, p_5, p_6, p_8) would be

zero. Furthermore, of the four possible non-zero probability functions, (p_1, p_3, p_7 and p_9), p_9 would be zero over the datasets associated with a valid learning hierarchy because the population value of P_{II} is then equal to zero; however, for the datasets associated with invalid hierarchies, p_9 is different from zero.

The error type "extremes" causes computer "memory overflow" due to abnormally large estimates occasionally obtained during the parameter estimation process of the MAX technique. Since the iterative parameter process involves complicated computations, such as the inversion of a 7x7 matrix at each iteration, this error type could be due to some computer rounding errors in the lengthy and complicated computations.

The two error types "SQRT5" and "SQRTP2" are both associated with computing the square root of the variance of $\hat{P}_{II}$ when this variance happened to be a negative value. For "SQRT5", convergence of estimates was checked at the end of the 5th iteration of the parameter estimation process; if convergence was detected, then the square root of the variance of $\hat{P}_{II}$ would be computed. "SQRTP2" is the same as "SQRT5" except that convergence of estimates was checked after each iteration rather than at the end of the 5th iteration.

The error types in category II are related to the problem of non-convergence of parameter estimates. These errors seemed to occur more often than the errors in category I and III.

Owston implemented a one-sided check for convergence of parameter estimates at the end of the 5th iteration of the parameter estimation process. This check, although inadequate, might still be used in detecting a large portion of the non-convergence cases. However, no non-convergence cases were reported by Owston. One possible explanation for this is that a simple but serious program logic error was committed by Owston(see Appendix E for details). The logical error, in effect, might lead Owston to consider a "non-convergent" case as "convergent".

By implementing a proper two-sided(using absolute value) check for convergence immediately after Owston's one-sided check, it was found that there were indeed some non-convergent cases that could not be detected by the one-sided check.

Since only 5 iterations were used by Owston, non-convergence of estimates appeared very likely. When ten iterations were used, there appeared to be a sizeable decrease in non-convergence cases; however, in some samples, the estimates never converged. It was identified that, in

these samples, the intermediate estimates usually had unreasonable values. Supposedly, the estimates should have values in the (0,1) range. Some of the intermediate estimates had values outside the (-1,2) range.

The error types in category III are associated with improperly converging estimates. It was found that some of the converging P_{II} estimates had negative and relatively large values(e.g. less than -.05). It was also identified that other estimates (there were six other estimates besides $\hat{P}_{II}$) might also be improperly converging. Since the MAX technique is mainly based on the P_{II} estimate, it appeared reasonable to emphasize only the proper convergence of $\hat{P}_{II}$; however, it may also be reasonable to consider the proper convergence of the other estimates at the same time because the estimates are interrelated in the process of estimation.

1.6 THE RESEARCH PROBLEM

According to Owston, the MAX technique appears to be a potentially very useful device for validating learning hierarchies. However, different types of errors existed in the computer program used by Owston to implement the MAX technique. This means that Owston's findings on MAX might be biased due to his failure to account for these errors.

It is therefore appropriate to raise the following questions about the MAX technique:

1. If all the errors were accounted for, would the MAX technique be useful for validating learning hierarchies?
2. If Owston's findings were seriously biased, is it possible to improve the procedure for validity decision making using the MAX technique?

Chapter II

DESIGN AND EXPERIMENTATION

At the end of the previous chapter, two questions were raised. The first question is related to the extent of error bias in Owston's data when using his computer program to implement the MAX technique. The second question is associated with the possible improvements on the implementation procedure of the MAX technique.

Corresponding to the two questions, two procedures, A and B, were developed. These two procedures are modified versions of Owston's procedure for implementing the MAX technique.

In this chapter, there are three major sections. The first section contains a description of procedures A and B. In the second section the design for collecting data by means of these two procedures is discussed. Suitable criteria to be used for analysing the data to be collected are reported in the final section.

2.1 TWO PROCEDURES FOR IMPLEMENTING THE MAX TECHNIQUE

Various types of error exist in Owston's computer program. Procedures A and B contain error checks that can be used to count the frequencies of occurrence of the different types of errors. There are similarities and differences between procedure A and procedure B . First, similarities are described.

Both procedures, A and B, are modified versions of Owston's procedure for implementing the MAX technique. These two procedures contain essential error checks that are mainly used to avoid computer program interruptions. Table V shows the various types of errors identified, together with the indications of in which procedure(s) each error type is being checked. For example, in error category I, errors related to "f=0", "p=0" and "extremes" cases are checked in both procedures. In fact, procedures A and B are designed in such a parallel manner that the three types of errors, just mentioned, are monitored by error checks which are common to the two procedures. Note that the two other error-types, in category I, are basically the same. They are related to the problem of computing the square root of a negative number. Two error checks are used, one for procedure A and the other one for procedure B, to check for "negative-square-root" cases.

Table V: Error types that are checked for by using Procedure A, Procedure B, or both procedures A and B

Error Category	Category Description	Error type No.	Error type Label	Procedure(s) used to check the error type.	Error Type Description
I	Error types in this category cause computer program interruptions. It is essential to check these types of errors.-	1	"f=0"	A , B	Marginal frequency f=0. Leads to "division by zero".
		2	"p=0"	A , B	probability values P1 , P2 , ... Pv, shown in Table I , approach zero.
		3	"Extremes"	A , B	Intermediate estimates take on extreme values.
		4	"SQRT5"	A	At the end of 5 iterations, variance of $\hat{P}_{II}$ may be a negative value. This cause computation of square root of a negative value.
		5	"SQRTP2"	B	It is the same as "-SQRT5" except that convergence is checked after each iteration instead of at the end of the 5th iteration.
II	Error types in this category are related to non-convergence of parameter estimates.	1	"NCONV5"	A	"Non convergence" of estimates is detected by an "inadequate"† check at the end of 5 iterations.
		2	"PNCONV5"	A	"Inadequate" check indicates "convergence" but proper†† check indicates non-convergence.
		3	"NCONV10"	B	Non-convergence of estimates at the end of 10 iterations.
		4	"INTVAL"	B	Intermediate estimates take on unreasonable values.
III	This category of errors are associated with improperly converged estimates	1	"-$\hat{P}_{II}$"	B	Converged $\hat{P}_{II}$ value happens to be a large negative value.
		2	"FINAL"	B	The final converged estimates, other than $\hat{P}_{II}$, are outside the normal (0,1) range.

† "inadequat" check for convergence is $|\Delta\Pi| < .001$, where Π is the vector of additive scores.

†† "proper" check is $|\Delta\Pi| < .001$

Since procedure A is used to collect data for determining the extent of bias in Owston's data, procedure A is set up to be as similar as possible to Owston's procedure. In procedure A, 5 iterations are used for the parameter estimation process. A check for convergence of estimates is performed at the end of the 5th iteration. In contrast, procedure B is used to provide possible improvements. In procedure B, a maximum of 10 iterations is allowed for the process of parameter estimation. Convergence check is made after each iteration so as to detect convergence of estimates which may occur between the first and the tenth iteration.

Another difference between the two procedures is the nature of the checks for convergence. Owston used an inadequate(one-sided; see Table V) check for convergence. This inadequate check is included in procedure A; it is followed by a proper(two-sided) check. The latter check detects cases that are non-convergent but that would be indicated as "convergent" by the inadequate check. In procedure B, only one proper check for convergence is utilized.

For the last error-type involving non-convergence of estimates, a conservative range check of (-1,2) is used to detect unreasonable intermediate parameter estimates outside this range. This range check is only used in procedure B.

Two other error-checks unique to procedure B, are utilized to detect improperly converging estimates. The first error-type in category III is related to a negative converging $\hat{P}_{II}$ value. A conservative check, $\hat{P}_{II}<-.05$, is used to determine if the value of $\hat{P}_{II}$ is negative and relatively large. The final error-check, unique to procedure B, is used to detect improperly converging estimates other than $\hat{P}_{II}$. A liberal range of (-.05, 1.05) is used to check for estimates (besides $\hat{P}_{II}$) having values outside this range.

Thus far, the similarities and the major differences between the two procedures, A and B, have been described. At the same time, all the important error checks used in these two procedures have been discussed. These error checks were implemented in a modified version of Owston's computer program. A computer flowchart showing the various error checks used in procedures A and B, appears in Appendix C.

Before proceeding on to the next section, it may be appropriate to mention another difference between procedures A and B. In procedure A, two levels are used in deriving confidence intervals around $\hat{P}_{II}$. The two levels are the 95% and 90% confidence intervals. In procedure B, besides these two levels, four other levels are used. They are the

85%,80%,75% and 70% confidence intervals. These additional levels are included because of Owston's suggestion that the MAX technique may perform better with shorter confidence intervals.

2.2 DESIGN FOR COLLECTING DATA

With minor modifications, Owston's design for studying the MAX technique was used. Modifications are required to accommodate the two procedures, A and B. At the same time, through modifications, the scope of the current study is reduced to a manageable size.

The random sample generation procedure for the current study is the same as Owston's. Owston used four levels of sample sizes of 25, 40, 100 and 400. The sample size of 25 appears to be inappropriate for the MAX technique because of the theoretical assumption of large sample size in the parameter estimation process. Also, it is suggested that the various types of identified errors occur frequently at sample sizes of 25. Consequently only three levels of sample sizes were used in this study. These levels are 40, 100 and 400.

In specifying population parameters, Owston used three levels of validity. The levels are valid, invalid with $\phi=0$, and invalid with $\phi=.25$. Owston later argued that two

levels would be sufficient. The level, invalid with $\phi=0$, is somewhat redundant. So, in this study, two levels of validity were adopted. The level invalid with $\phi=0$ was eliminated.

Besides validity of the hierarchy, one other factor, used by Owston in parameter specification, is skill difficulty. From Owston's four levels of skill difficulty pairs, (D_I, D_{II}), where D_I and D_{II} are skill difficulties of skill I and skill II respectively, three levels were chosen. These three levels are (.95, .90), (.70, .50) and (.40, .25). The level not adopted was related to extremely difficulty skills with (D_I, D_{II}) being (.10, .05); the reason for not adopting this level is that it might not be frequently encountered in practice.

In the White & Clark Model of learning hierarchies, there are four population parameters related to four mutually exclusive groups of subjects with different status of skill possession. These four parameters P_O, P_I, P_{II} and P_B correspond to the four proportions of subjects who possess "neither skill", only skill I, only skill II, and "both skill I and skill II" respectively.

When three levels of skill difficuty, (D_I, D_{II}), are combined with two levels of validity, six sets of (P_O, P_I, P_{II}, P_B) values can be determined by following the procedure

used by Owston. The six sets of P values are shown in Table VI.

Besides the four P parameters, there are four other parmeters, in the White & Clark Model, related to the probabilities of correctly answering skill I and skill II test items. Owston allowed the probabilities of correctly answering the items through guessing to vary at four levels. These four levels were .25, .20, .10 and .05 . In Owston's discussions, .25 and .20 levels would be related to the guessing probabilities of multiple choice types of items with four and five alternatives respectively. The two levels, .10 and .05, would be related to guessing probabilities of short-answer types of items. Three of the four levels of probabilities of guessing were used in this study. These three levels are .25, .20, .05 .

By combining six sets of P values and 3 levels of probabilities of guessing, 18 sets of population parameters were formed. Using each of the 18 sets of population parameters random samples of three different sample sizes were generated. The three sample sizes were 40,100 and 400. At each level of sample size, for each set of population parameters, 1000 random samples were generated. Each "1000 samples" formed a dataset for testing the MAX technique via procedures A and B.

Table VI: Six sets of P's corresponding to 3 levels of skill difficulty and 2 levels of validity of learning hierarchy

(Diff. skill I, Diff. skill II) (D_I, D_{II})	Valid	Invalid
	(P_0, P_I, P_{II}, P_B)	(P_0, P_I, P_{II}, P_B)
(.95, .90)	(0.05, 0.05, 0.00, 0.90)	(0.02, 0.08, 0.03, 0.87)
(.70, .50)	(0.30, 0.20, 0.00, 0.50)	(0.21, 0.29, 0.09, 0.41)
(.40, .25)	(0.60, 0.15, 0.00, 0.25)	(0.50, 0.25, 0.10, 0.15)

*Each set of P's contains four values representing four mutually exclusive proportions of subjects with different status of skill possession.

By combining 18 sets of parameters with 3 sample sizes, 54 sets of 1000 random samples were drawn, 27 of these datasets were related to valid hierarchies at 3 different levels of skill difficulty, probability of guessing and sample size. The other 27 datasets were related to invalid hierarchies with the corresponding levels of skill difficulty, probability of guessing, and sample size.

In Owston's study, for each dataset, his procedure was applied to make 1000 validity decisions using 95% confidence intervals and to make another 1000 validity decisions using 90% confidence intervals. However, due to the existence of the various types of errors, some modifications were made in this part of his design. First, assume that in each dataset x samples are detected by procedure A to be associated with the different error-types; correspondingly, y samples are detected by procedure B. Note that x and y are variables taking on different values over the different datasets. Then, the number of samples, in each dataset, that are used to make decisions in procedure A, would be "1000-x". Correspondingly, the number of samples, used to make decisions, in procedure B, would be "1000-y".

For procedure A, two confidence intervals (C.I.), 95% and 90%, were used. "1000-x" decisions were therefore made at 95% confidence interval, and another "1000-x" decisions at 90% confidence intervals.

For procedure B, six levels of confidence intervals, 95%, 90%, 85%, 80%, 75% and 70%, were used. "1000-y" validity decisions were made at each of the six levels of confidence intervals.

As a summary of this section, a schematical representation of the design for the current study is shown in Table VII. The technical details of this design can be found in Appendix D, where the related computer program statements are listed.

2.3 CRITERIA FOR ANALYSING THE DATA

Validity decisions may be correct or incorrect. However, before applying procedures A and B to make validity decisions, the presence of errors was checked.

Samples, detected in procedure A as containing error, were not used for making validity decisions. Then, procedure A was, in effect, applied to the remaining samples to make decisions at two levels of confidence intervals. At each level of confidence interval, the results were two proportions(out of 1000), one proportion being related to correct decisions, and another proportion being related to incorrect decisions. Hence, in a dataset of 1000 random samples to test the MAX technique for procedure A, there was a constant proportion, P(ND), of samples related to "no-

Table VII

Schematical Representation of the design for applying procedures A and B to test the SII technique over different decision criteria, sample sizes, measurement errors, skill difficulty levels and validity levels

18 sets of population parameters are formed by the following combinations:			Based on each set of parameter values, random samples are generated. 1000 samples are generated at each of the following 3 levels of sample size:	Two procedures, A and B, are applied to the 1000 samples generated, one sample at a time. A and B have some different error checks, so the two procedures may have different error counts. Furthermore A and B do not have same levels of decision criteria. The set-ups for A and B are as follows:			
2 levels of validity	3 levels of skill difficulty (skill I, skill II) or. (D_I, D_{II})	3 levels of measurement errors of guessing		procedure	# of samples with error(s)	# of samples used for decision making	levels of decision criteria used on each sample involved in decision making
				A	x	1000-x	95% C.I. *
							90% C.I.
				B	y	1000-y	95% C.I.
	(0.95, 0.90)	0.05	40				90% C.I.
valid							85% C.I.
	(0.70, 0.50)	0.20	100				80% C.I.
invalid							75% C.I.
(e=0.25)	(0.40, 0.25)	0.25	400				70% C.I.

* C.I. indicates confidence interval

decision" because of errors. The remaining proportion, 1-P(ND), was divided into a proportion of correct decisions, P(COR), and a proportion of incorrect decisions, P(INC). P(COR) and P(INC) could vary from the 95% C.I. to the 90% C.I. level, but the sum of P(COR) and P(INC) was a constant for that particular dataset.

Similarly, procedure B was applied in a like fashion, except that with procedure B, four additional levels of decision criteria were used in constructing confidence intervals.

Owston used the proportion of incorrect decisions as an estimate of the risk of committing decision errors. In view of the additional proportion of "no-decision" cases, it appeared to be more useful to adopt the proportion of correct decisions as a measure of the performance of the MAX technique.

Corresponding to Owston's criteria of requiring the proportion of incorrect decisions to be less than or equal to .20 for valid hierarchies, and less than or equal to .05 for invalid hierarchies, proportion of correct decisions, P(COR) was considered as satisfactory if it was greater than .80 for valid hierarchies, and .95 for invalid hierarchies.

The (.80, .95) criteria, just mentioned, will be

utilized, in the next chapter, as a basis for comparing the current results with those found by Owston.

Chapter III

PRESENTATION AND DISCUSSIONS OF RESULTS

At the end of the first chapter, two research questions were raised. The first question is related to determining the extent to which the various types of identified errors affect the performance of the MAX technique for validating learning hierarchies. The second question is about the possibilities of improving the procedure for implementing the MAX technique.

Corresponding to these two questions, two procedures, A and B, were designed. Procedure A is identical to Owston's procedure for implementing the MAX technique, except that, in procedure A checks were provided to detect the various types of identified errors. Procedure B was designed to improve the performance of the MAX technique; additional error checks and modified decision criteria were utilized in procedure B.

Procedures A and B were applied to 27 datasets characterising valid learning hierarchies(for convenience, these datasets are called valid datasets) and to 27 corresponding datasets characterising invalid

hierarchies(for convenience, these datasets are called invalid datasets). Corresponding to each type of identified errors relative frequencies of occurrence were calculated when applying procedures A and B. Also, proportions of correct and of incorrect decisions at each level of decision criteria were calculated.

In this chapter, results obtained from using procedure A and B, over the 27 valid and the 27 invalid datasets, are presented and discussed. For the purpose of comparison, pertinent results from Owston's study are also included.

3.1 RESULTS OBTAINED FROM USING PROCEDURE A

The results, obtained from using procedure A, were divided into two parts. The first part is related to the various error types. The second part is related to the proportions of correct and incorrect validity decisions. Since procedure A was designed mainly for investigating the extent of influences due to the identified errors, the emphasis is placed on the first part of the results. In the next section, the proportions of "correct decisions" obtained by using procedure A are compared with the corresponding findings in Owston's study.

In Table VIII, the frequencies of occurrence of the various types of errors in each different dataset (1000

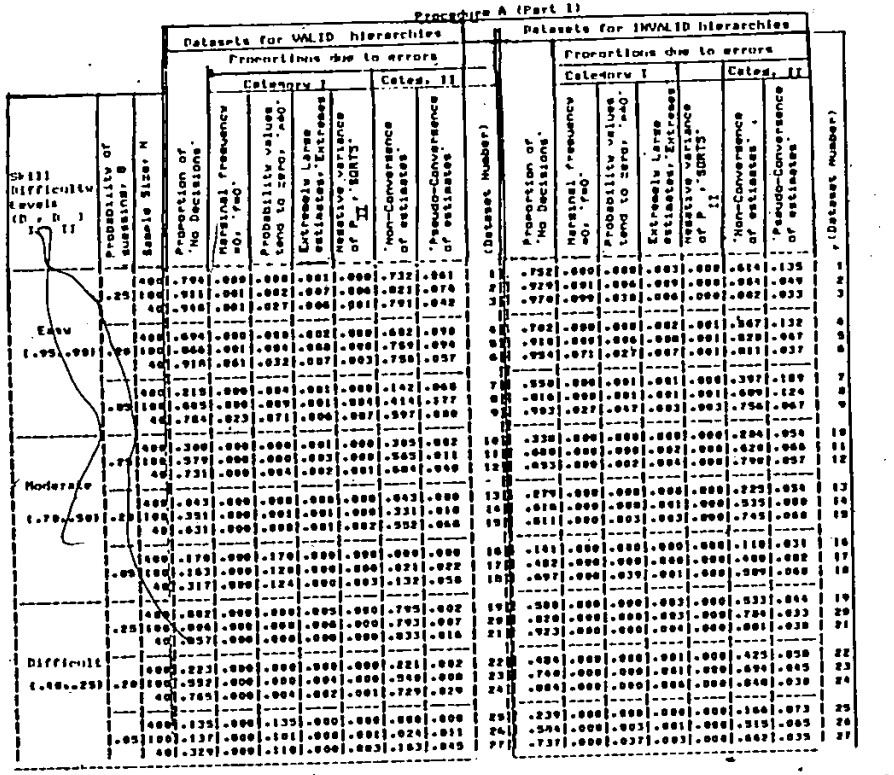


Table VIII: Proportions of the various types of errors detected by using procedure A, for the different levels of validity, split difficulty, probability of success and sample size.

Notes:
1. Procedure A is the same as Owston's procedure for using the RAI technique to validate learning hierarchies except that error checks are available in procedure A.
2. For each dataset, the proportion of "no decisions" is the sum of the proportions due to the various types of errors.

samples per dataset) are shown as proportions. For each dataset, the sum of the proportions due to the six types of errors ("f=0", p=0,...etc.) is denoted as the proportion of "no decisions". In general, the proportions of "no decisions" are fairly high; over the valid datasets, the range is from .043 to .948 and over the invalid datasets, the range is from .141 to .970 . For each dataset, a high proportion of "no decisions" implies that there would be a small number of samples for which validity decisions (correct or incorrect decisions) could be made. Also, it implies that some proportions related to the various types (6 types) of identified errors are high. Recall that these six types of errors belong to two error categories. The first four of the six error types belong to error category I which is related to errors that may cause computer interruptions. The last two error types, "non-convergence" and "pseudo-convergence", belong to error category II which is related to the problems of non-convergence of parameter estimates.

It may be observed, in Table VIII, that the proportions related to the error types in category I are generally low as compared with the proportions related to the error types in category II. This implies that, in general, computer interruptions would not occur very frequently. Furthermore, it implies that, using procedure A, non-

convergence of parameter estimates occurs frequently. Since, in procedure A only five iterations were allowed for the parameter estimation process, it is not surprising to observe high proportions of category II error types.

It may be interesting to note, that for each error type the patterns of the proportions associated with the valid datasets may be very different from the corresponding patterns of the proportions associated with the invalid datasets. For example, for the error type related to very small probability values ($p\doteq 0$), the proportions in some valid datasets (e.g. dataset # 16,17,18,25,26,27) were in general much higher than in the invalid counterparts. The explanation of this phenomenon is postponed until each error type is studied in more detail. Another example can be used to illustrate a somewhat reversed pattern in which the proportions related to invalid datasets are higher than those from the valid counterparts. For the error type, "non-convergence" of estimates, it is generally true that the proportions in the invalid datasets are relatively higher than the corresponding proportions in the valid datasets. The contrast is especially clear in the same datasets (# 16,17,18,25,26,27) as when the error type "$p\doteq 0$" was found to occur more often over the valid hierarchies.

In the subsequent paragraphs, each error type with the

two error categories will be studied more carefully. The incidence of each type of error will be compared for valid or invalid hierarchies. Frequent references will be made to Table VIII where the error proportions are shown. For the detailed descriptions of each error type, the reader is referred to sections 1.5 and 2.1 .

The error type "marginal frequency equals zero" may occur in the computation of initial parameter estimates. One of the marginal frequencies, f, used in calculating the initial estimates may happen to be zero. If that were the case, computer interruption would occur due to an undefined computational situation of dividing some values by zero. An error check was used to detect this error type and is used to by-pass computer interruptions.

As shown in Table VIII, this error type was detected in datasets associated only with easy skills learning hierarchies and small sample sizes (mainly sample sizes of 40). Probability of guessing appears to be related to the size of the proportions due to this type of error. As the probability of guessing decreases, the magnitudes of the error proportions also decrease. These patterns were expected to occur. Reasons for such expectations were given in the part of section 1.5 where this error type was discussed in detail. Also, it may be noticed that the

proportions associated with this error type are relatively small. Over the valid datasets, the range is from .000 to .081 . Over the invalid datasets, the range is from .000 to .099 .

The error type "probability values tend to be zero" may occur during the parameter estimation process when the probability functions in the White & Clark Model(shown in Table I) are used to compute the first approximation of the parameter estimates. If one or more of these probability functions approach zero, computer interruption may result from a near zero denominator in some expressions. In Table VIII, proportions due to this type of error are found to be generally higher in the valid datasets than in the invalid datasets. The range, over the valid datasets, is from .000 to .170 . The range, over the invalid datasets, is from .000 to .047 . It is particularly interesting to note that the major differences in proportions, between valid and invalid datasets, due to this error type, occurred in the datasets related to moderate and difficult skills hierarchies with a probability of guessing level of .05 .

One possible explanation for the above phenomenon is that the probability functions(shown in Table I) are different over the valid and invalid datasets. For example, assume that there were no measurement errors(such that the

probability of guessing is zero, etc.), one of the probability functions(P_{O2}; shown in Table I) is in fact equal to zero in the valid datasets; however this probability function is different from zero in the invalid datasets.

The error type "extremely large estimates" may occur during the parameter estimation process. Due to some unknown reasons (most likely due to the complicated matrix inversion process involved in the MAX technique), the intermediate parameter estimates may(although rarely) be abnormally large (larger than ten to the power 6). An error check was used to detect this error type so as to avoid computer interruptions due to "computer memory overflow". As shown in Table VIII, the proportions associated with this error type are of small magnitude. Over the valid datasets, the range of proportions, due to this error type, is from .000 to .008 . Over the invalid datasets, the range is from .000 to .009 .

The error type "negative variance of $\hat{P}_{II}$" could occur when the confidence interval around $\hat{P}_{II}$ is constructed. Basically, this error type could be due to rounding errors in the complicated computational processes involved in the MAX technique. However, it may be observed, in Table VIII, that this type of error affected the valid datasets slightly

more often than the invalid datasets. Over the valid datasets, the proportions over this error type range from .000 to .007; while over the invalid datasets, the range is from .000 to .003 . Particularly interesting to note is that this error type was never detected in the invalid datasets associated with moderate and difficult skills hierarchies. The real reasons behind this phenomenon are uncertain. Additional data may be useful for further understanding. However, one possible explanation is that the variance estimate of P_{II} is directly related to the estimated value of P_{II} which may be negative due to the normal approximation of the distribution of the parameters around zero. For the valid datasets, the parameter P_{II} was set at zero; as for the invalid datasets, P_{II} was larger than zero. In the valid datasets, it is more likely to have negative estimates of P_{II}; this may lead to the more frequent occurrence of negative variance estimate of P_{II} in the valid datasets.

So far, each of the error types, in error category I, which can be detected when using procedure A, has been discussed. Also mention was made of the "non-convergence" error type in error category II. The other error type, "pseudo convergence of estimate", in category II, may be studied together with the "non-convergence" type because these two error types are both related to the problem of

non-convergence of parameter estimate. The reason for designing a separate error check was to investigate the deficiencies of Owston's check (one-sided check on additive corrections; the proper check should be based on absolute values of the additive corrections for testing the convergence condition, e.g. "|additive corrections|<.001").

In general, the error types in category II occur more often over the invalid datasets. A possible explanation of this phenomenon is that the initial parameter estimate values of P_{II} (the proportion of subjects possessing skill II without possessing skill I) were set to be zero for both valid and invalid datasets. For the valid datasets, the population parameter values of P_{II} were in fact zero; however, for the invalid datasets, the corresponding values were greater than zero; for the easy, moderate and difficult skills hierarchies, the P_{II} values for invalid datasets were, respectively .03, .09 and .10 . It would be reasonable to expect that, for the invalid datasets, convergence of the parameter estimate of P_{II} from zero to the population values (.03, .09 and .10) was more difficult than the convergence of P_{II} estimate value from zero to zero in the valid datasets.

Besides validity of the hierarchy, other factors such as sample sizes, probability of guessing levels and skill

difficulty levels also appeared to affect the error types in category II. For datasets based on large sample sizes of 400, the error proportions were generally lower than the corresponding proportions related to sample sizes of 100 and 40. The lower error proportions also tend to occur with the low measurement error for guessing. Furthermore, the error proportions, in the datasets associated with moderate skills learning hierarchies, were generally lower than those from the difficult and easy skill learning hierarchies. This phenomenon may be due to the fact that the difference, $D_I - D_{II}$ (where D_I, D_{II} are the difficulty levels of skill I and skill II respectively), is larger over the moderate skill learning hierarchies. For the moderate skills learning hierarchies, $D_I - D_{II}$ is 0.20; for the difficult skills hierarchies, $D_I - D_{II}$ is 0.15; while, for the easy skills, $D_I - D_{II}$ is 0.05 .

So far, only the error proportions have been discussed. The rest of this section is used to present and discuss the proportions of correct and incorrect decisions.

As shown in Table IX, over the valid datasets, the proportions of correct decisions tend to be smaller at the narrower 90% confidence interval. However, over the invalid datasets, the proportions of correct decisions tend to be larger at the 90% confidence interval. According to the

Procedure A (Part II)

Skill Difficulty Levels (D_I, D_{II})	Probability of Guessing, g	Sample Size, N	Datasets for VALID hierarchies: Proportions of 'No Decisions'	95% Confidence Interval: Correct Decisions	95% Confidence Interval: Incorrect Decisions	90% Confidence Interval: Correct Decisions	90% Confidence Interval: Incorrect Decisions	Datasets for INVALID hierarchies: Proportions of 'No Decisions'	95% Confidence Interval: Correct Decisions	95% Confidence Interval: Incorrect Decisions	90% Confidence Interval: Correct Decisions	90% Confidence Interval: Incorrect Decisions
Easy (.95..90)	.25	400	.794	.206	.000	.205	.001	.752	.001	.247	.028	.220
		100	.911	.089	.000	.089	.000	.929	.000	.071	.000	.071
		40	.948	.051	.001	.051	.001	.970	.006	.024	.009	.021
	.20	400	.694	.306	.000	.304	.002	.702	.021	.277	.066	.232
		100	.866	.134	.000	.134	.000	.910	.001	.089	.001	.089
		40	.918	.080	.002	.080	.002	.954	.008	.038	.013	.033
	.05	400	.215	.785	.000	.782	.003	.558	.217	.225	.326	.116
		100	.605	.395	.000	.395	.000	.816	.006	.178	.017	.167
		40	.786	.215	.001	.214	.002	.903	.018	.079	.026	.071
Moderate (.70..50)	.25	400	.308	.690	.002	.682	.010	.338	.309	.353	.402	.180
		100	.579	.421	.000	.421	.000	.688	.003	.309	.013	.299
		40	.731	.269	.000	.269	.000	.853	.000	.147	.000	.147
	.20	400	.043	.953	.004	.944	.013	.279	.372	.149	.663	.058
		100	.351	.649	.000	.647	.002	.616	.013	.371	.048	.336
		40	.631	.369	.000	.369	.000	.811	.001	.188	.002	.187
	.05	400	.170	.830	.000	.830	.000	.141	.859	.000	.859	.000
		100	.163	.837	.000	.836	.001	.482	.257	.261	.398	.120
		40	.317	.683	.000	.683	.000	.697	.006	.297	.029	.274
Difficult (.40..25)	.25	400	.802	.195	.003	.190	.008	.580	.093	.327	.195	.225
		100	.806	.194	.000	.194	.000	.820	.000	.180	.006	.174
		40	.857	.143	.000	.143	.000	.923	.000	.077	.000	.077
	.20	400	.223	.772	.005	.761	.016	.486	.282	.234	.406	.112
		100	.552	.448	.000	.446	.002	.740	.004	.256	.018	.242
		40	.765	.235	.000	.234	.001	.884	.000	.116	.002	.114
	.05	400	.135	.865	.000	.864	.001	.239	.761	.000	.761	.000
		100	.137	.863	.000	.863	.000	.584	.225	.191	.336	.080
		40	.329	.671	.000	.671	.000	.737	.011	.252	.047	.218

Table III: Proportions of correct decisions and incorrect decisions at 95% and 90% confidence intervals, made by using procedure A, over the various combinations of validity, skill difficulty, probability of guessing and sample size levels.

Notes: The proportions of 'No Decisions' are the same as the corresponding entries in Table VIII.

nature of the decision process involved in the MAX technique, the above phenomenon is logical. The narrower 90% confidence interval (as compared with the 95% confidence interval) is less likely to contain zero. Hence, the probability of indicating any hierarchy as a valid hierarchy decreases with the decrease in the width of the confidence interval. Consequently, for a valid dataset, the proportion of correct decisions tends to decrease with the decrease in the width of confidence interval. In contrast , for an invalid dataset, the proportion of correct decisions tends to increase with the decrease in the width of the confidence interval; because it is less likely to contain zero.

It may also be observed that the proportions of incorrect decisions are comparatively smaller over the valid datasets than over the invalid datasets. For example, at the 95% confidence interval, the range of incorrect decisions, over the valid datasets, is from .000 to .005 . Over the invalid datasets the range is from .000 to .371 .

This phenomenon might be related to the values of the P_{II} estimates upon which confidence intervals were built and validity decisions were made. Supposedly, the convergent $\hat{P}_{II}$ estimate values would be very close to the population values. For the valid datasets, the population values of P_{II} are all equal to zero. For the invalid datasets, the

population values of P_{II} depend on the skill difficulty levels . For the datasets associated with the easy, moderate and difficult skills learning hierarchies, the population P_{II} values are .03, .09 and .10 respectively. These values are distinctly different from zero.

For the invalid datasets, if the estimated values of P_{II} were very close to the corresponding population values, the confidence intervals built around P_{II} estimates should not likely contain zero. Hence, there should be little chance of incorrectly indicating an invalid hierarchy as valid. However, the high proportions of incorrect decisions, over the invalid datasets, indicate otherwise.

Recall that, in the iterative parameter estimation process used in the MAX technique, an initial set of parameter estimates is required. The White & Clark marginal frequency method for obtaining parameter estimates(from subjects' response patterns to two skill testing items per skill) was used. In this method the initial estimate values for P_{II} are set to zero for all the valid and invalid datasets. Also, recall that in discussing the problem of non-convergence of parameter estimates, it was mentioned that it might not be easy to obtain the convergence of $\hat{P}_{II}$ in the invalid datasets.

Another possibility for the occurrence of high proportions of incorrect decisions, over the invalid datasets, is that the variance estimates of P_{II} used to construct confidence intervals were so high that the chances of including zero in the confidence intervals were also high. This also might lead to the high proportions of incorrect decisions over the invalid datasets.

3.2 COMPARISONS OF OWSTON'S RESULTS WITH THE RESULTS OBTAINED BY USING PROCEDURE A

At the end of chapter II, criteria were established for comparing results. For the valid datasets, the proportion of "correct decisions" was required to be larger than .80 before the MAX technique was considered as satisfactory. For the invalid datasets, the proportion of "correct decisions" was required to be larger than .95 . Also, these criteria were required to be met simultaneously in the valid datasets and in the corresponding (having the same level of skill difficulty, probability of guessing and sample size) invalid datasets.

Based on these criteria, (.80, .95), Owston's results (retrieved from Appendix of Owston's thesis), and the proportions of correct decisions obtained from using procedure A (retrieved from Table IX) were compared. The data used for comparisons are shown in Table X.

Skill Difficulty Levels (P_I, P_{II})	Probability of guessing, θ	Sample Size, N	Proportions of "correct Decisions"											
			95% Confidence Interval						90% Confidence Interval					
			Duston'Results		Results from Procedure A		Adjusted Results from Procedure A		Duston's Results		Results from Procedure A		Adjusted Results from Procedure A	
			Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid
Easy (.95,.90)	.25	400	0.934	0.809	0.706	0.001	0.206	0.753	0.923	0.825	0.205	0.028	0.205	0.780
		100	0.920	0.693	0.089	0.000	0.097	0.923	0.918	0.708	0.089	0.000	0.097	0.923
		40	0.951	0.671	0.051	0.006	0.079	0.946	0.941	0.633	0.051	0.009	0.079	0.949
	.20	400	0.943	*****	0.306	0.021	0.306	0.755	0.959	*****	0.304	0.086	0.304	0.767
		100	0.944	0.706	0.134	0.001	0.138	0.904	0.934	0.732	0.134	0.001	0.138	0.904
		40	0.945	0.627	0.080	0.008	0.115	0.934	0.938	0.646	0.080	0.011	0.115	0.932
	.05	400	0.999 *	0.957	0.785	0.217	0.789	0.774	0.998 *	0.959	0.782	0.326	0.784	0.883
		100	0.982	0.811	0.395	0.006	0.408	0.820	0.974	0.831	0.395	0.017	0.408	0.831
		40	0.970	0.692	0.215	0.018	0.293	0.871	0.963	0.705	0.214	0.026	0.282	0.877
Moderate (.70,.50)	.25	400	0.979 *	0.993	0.690	0.309	0.690	0.647	0.969 *	0.994	0.682	0.482	0.682	0.870
		100	0.946	0.828	0.421	0.003	0.421	0.691	0.934	0.853	0.421	0.013	0.421	0.701
		40	0.937	0.649	0.269	0.000	0.274	0.851	0.934	0.629	0.269	0.000	0.274	0.851
	.20	400	0.995 *	0.998	0.953	0.572	0.953	0.851	0.985 *	0.999	0.944	0.663	0.944	0.942
		100	0.981	0.885	0.649	0.013	0.650	0.630	0.978	0.903	0.647	0.048	0.648	0.665
		40	0.979	0.712	0.369	0.001	0.379	0.809	0.974	0.744	0.369	0.002	0.379	0.810
	.05	400	1.000 *	1.000	0.830	0.859	1.000 *	1.000	0.997 *	1.000	0.810	0.859	1.000 *	1.000
		100	1.000 *	0.990	0.837	0.257	0.957	0.739	1.000 *	0.991	0.836	0.398	0.956	0.880
		40	1.000	0.911	0.683	0.006	0.810	0.664	0.999	0.914	0.683	0.029	0.810	0.687
Difficult (.40,.25)	.25	400	0.993	0.878	0.195	0.693	0.195	0.673	0.977	0.924	0.190	0.195	0.190	0.775
		100	1.000	0.772	0.194	0.000	0.194	0.820	0.991	0.776	0.194	0.004	0.194	0.856
		40	0.994	0.728	0.143	0.000	0.143	0.923	0.991	0.731	0.143	0.000	0.143	0.923
	.20	400	0.993 *	0.992	0.777	0.282	0.772	0.744	0.979 *	0.992	0.761	0.404	0.761	0.808
		100	1.000	0.868	0.448	0.004	0.448	0.744	0.997	0.876	0.446	0.018	0.446	0.758
		40	0.998	0.686	0.235	0.000	0.240	0.884	0.996	0.722	0.234	0.002	0.239	0.886
	.05	400	1.000 *	0.999	0.845	0.761	1.000 *	1.000	0.990 *	0.999	0.844	0.761	0.999 *	1.000
		100	1.000 *	0.970	0.843	0.275	0.965	0.806	0.999 *	0.972	0.843	0.334	0.965	0.917
		40	1.000	0.909	0.671	0.011	0.792	0.731	1.000	0.923	0.671	0.047	0.792	0.747

Table 1: Comparisons of the proportions of "correct decisions", at decision criteria of 95% and 90% confidence intervals, among results obtained from three sources (Duston's results; Results from Procedure A; Results from Procedure A with adjustments to include some "No decisions" cases as "correct decisions" cases), over the various combinations of validity (valid & invalid), skill difficulty, probability of guessing, and sample size levels.

* Indicates that the two criteria, requiring the proportions of "correct decisions" to be larger than .80 for valid hierarchies and to be larger than .90 for invalid hierarchies, were simultaneously satisfied.

***** Indicates that data were missing from Duston's results.

It may be noticed that some extra data grouped under the headings of "Adjusted Results from Procedure A" are included. The reasons for these adjustments will be explained in the next few paragraphs.

As discussed in the last section, different error types may have different effects on the valid and invalid datasets. Thus, it appears possible to utilize this property of the error types to aid in making validity decisions.

As an extreme case, suppose that a certain error type occurred only in the valid datasets and never occurred in the invalid datasets. Further, suppose that there were sufficient evidence to believe that such a pattern was reasonable. If, based on this information, the researcher treated the occurrences of this type of error as an indication of a valid hierarchy, he would increase the proportions of correct decisions over the valid datasets by the corresponding proportions due to this type of error. At the same time, the proportions of correct decisions over the invalid datasets would not be affected by such adjustments over the valid datasets. However, the proportions of incorrect decisions over the invalid datasets would be increased by the proportions related to this error type.

Similarly, an error type that was detected mainly in the invalid datasets, might be treated as an indication of an invalid hierarchy. Consequently, the proportions of correct decisions over the invalid datsets would be increased without affecting the proportions of correct decisions over the valid datasets. Nevertheless, the proportions of incorrect decisions over the valid datasets would be increased.

Assume that, there is another error type which was detected as often among the valid datasets as among the invalid datasets. Suppose this error type had to be classified as an indication of a valid or an invalid hierarchy. The logical action is to treat this type of error as an indication of an invalid hierarchy because, from a conservative viewpoint, it is a less serious decision error to consider a valid learning hierarchy as invalid than to consider an invalid learning hierarchy as valid.

If the information on the relative incidences of proportions of the various types of errors were not fully utilized, there would be sizeable proportions of "no decisions" due to the various error types. However, the information might be profitably used to eliminate the incidences of no decisions.

Recall that two error types, "negative variance of $\hat{P}_{II}$" and "probability values tend to be zero", discussed in the last section, were detected more often over the valid datasets. There was evidence for believing that such patterns were reasonable. If each of these two types of errors is treated as an indication of a valid learning hierarchy, the proportion of "correct decisions" over each of the valid datasets can therefore be adjusted upward to include the proportions corresponding to the above two error types. This results in little change in the error rates for invalid hierarchies.

The error types, "marginal frequency equals zero" and "extremely large parameter estimates" appear to be detected as often among the valid datasets as among the invalid datasets. As discussed earlier, based on a conservative viewpoint (considering that it is a less serious error to indicate a valid hierarchy as invalid than to indicate an invalid hierarchy as valid), each of these two error types is therefore treated as an indication of an invalid hierarchy. The proportion of "correct decisions" over each of the invalid datasets is increased to include the proportions due to these two error types. These changes over the invalid datasets result in only slight changes in the error rates for the valid hierarchies. For each of the valid datasets, the proportion of "incorrect decisions" is

increased slightly by the inclusion of the proportions due to these two error types.

The other two error types, "non-convergence of estimates" and "pseudo-convergence of estimates", were generally detected more often over the invalid datasets. So, each of these two error types is treated as an indication of an invalid hierarchy. However, it may be observed that, in so doing, the proportion of correct decisions in each of the invalid datasets is increased sizeably (in general). The reason is that in each dataset, the proportion due to each of these two error types is fairly large. At the same time, the proportion of incorrect decisions in each of the valid datasets is adjusted upward to include the proportions due to these two error types.

The adjusting process, just discussed, seems to be reasonable, because the researcher is willing to increase the proportions of correct decisions over one type of learning hierarchies (e.g. invalid hierarchies) at the expense of increasing the proportions of incorrect decisions over the other type of learning hierarchies (e.g. valid hierarchies).

The adjusted results, as discussed above are shown in Table X. However, it may be appropriate to stress here that the adjusted results cannot be considered (in the strict

these) as directly obtained from applying the MAX technique. The reason is that only some of the samples (not all) were involved in the complete process of parameter estimation, confidence interval construction and decision making. However, from a practical point of view, by treating each error type globally as an indication of a valid or an invalid hierarchy, only two categories of proportions, "correct" and "incorrect" were possible.

In fact, the adjusted results may be considered as obtained from a "hybrid" procedure for validating learning hierarchy. The reason for calling it a "hybrid" procedure is that this procedure is based partly on the MAX technique and partly on the proportions of errors detected in applying the MAX technique. Strictly speaking, only the unadjusted proportions of correct decisions may be used for comparing with the corresponding results obtained by Owston.

Using data obtained earlier (see Table X) Owston would claim that the (.80, .95) criteria, mentioned at the beginning of this section, were met 8 times at each of the 95% and 90% confidence intervals. However, from the unadjusted results obtained from using procedure A, the (.80, .95) criteria were not met for any of the same 27 valid and 27 invalid datasets. With the adjusted results, the (.80, .95) criteria were satisfied twice at each of the

95% and 90% confidence intervals. The (.80, .95) criteria were met in the two pairs (each pair contains one valid and one invalid dataset) of datasets related to moderate and difficult skills with probability of guessing of .05 and sample sizes of 400.

From the above comparisons, it can be concluded that Owston's results are different from the results obtained by using procedure A (which is identical to Owston's procedure for data collection with the exception that procedure A contains error checks) before and after the adjustments were made. This conclusion leads the researcher to question the findings obtained by Owston over the datasets utilized in the current study.

In the next section, the results obtained from using procedure B will be presented and discussed.

3.3 RESULTS OBTAINED FROM USING PROCEDURE B

As mentioned before, procedure B was designed as an improved version of Owston's procedure for validating learning hierarchies. In procedure B, some additional error checks were used. Also, there were four extra levels of decision criteria of 85%, 80%, 75% and 70% confidence intervals. Ten iterations rather than five were used in the parameter estimation process.

The results obtained from using procedure B are divided into two parts. The first part (shown in Table XI) is related to the various types of identified errors. There are altogether eight error types (classified into three error categories). Detailed descriptions of each type of error may be found in sections 1.5 and 2.1 . The second part (shown in Table XII) of the results is related to the proportions of correct and incorrect decisions at each of the 95%, 90%, 85%, 80%, 75% and 70% confidence intervals.

In studying the first part of the results, the effects (on the various error proportions) of validity, skill difficulty, probability of guessing and sample sizes will be considered.

In Table XI, the proportions related to the various error types are shown. There are altogether three categories of error types detected in procedure B. The category I error types include "marginal frequency equals zero", "probability values tend to be zero", "extremely large estimates", and "negative variance of $\hat{P}_{II}$". It may be recognized that these error types were studied in the discussions of the results obtained from procedure A. In fact, the first three of the above four error types are identical to the corresponding ones detected by using procedure A because, for those error types , the checks used were common to both procedure A and procedure B.

Procedure B (Part I)

Skill difficulty levels (P_1, P_{11})	Probability of guessing: B	Sample Size: N	Datasets for VALID hierarchies: Proportion of "No Decisions"	Proportions due to errors: Category I: Marginal frequency =0, "P=0"	Category I: Probability values tend to zero, "P=0"	Category I: Extremely Large estimates, "Extremes"	Category I: Negative variance of P_{11}, "SORTP2"	Categ. II: Non-convergence of estimates, "NCONV10"	Categ. II: Unreasonably large estimates, "INTVAL"	Categ. III: Converging P_{11} being a large negative value	Categ. III: Other conversions estimates being outside (0,1) range.	Datasets for INVALID hierarchies: Proportion of "No Decisions"	Proportions due to errors: Category I: Marginal frequency =0, "P=0"	Category I: Probability values tend to zero, "P=0"	Category I: Extremely Large estimates, "Extremes"	Category I: Negative variance of P_{11}, "SORTP2"	Categ. II: Non-convergence of estimates, "NCONV10"	Categ. II: Unreasonably large estimates, "INTVAL"	Categ. III: Converging P_{11} being a large negative value	Categ. III: Other conversions estimates being outside (0,1) range.
Easy (.95..99)	.25	400	.385	.000	.000	.001	.000	.001	.328	.000	.005	.543	.000	.000	.003	.000	.099	.298	.000	.143
		100	.784	.001	.002	.007	.000	.007	.620	.000	.057	.871	.001	.000	.009	.000	.001	.678	.000	.098
		40	.883	.001	.027	.006	.000	.056	.682	.003	.026	.899	.099	.030	.006	.001	.050	.668	.000	.045
	.20	400	.261	.000	.000	.002	.000	.022	.222	.000	.005	.493	.000	.000	.002	.000	.000	.241	.000	.152
		100	.697	.001	.004	.000	.000	.004	.533	.000	.007	.843	.000	.006	.000	.000	.010	.501	.000	.130
		40	.847	.001	.032	.007	.002	.072	.631	.002	.040	.898	.071	.027	.007	.001	.051	.664	.001	.048
	.05	400	.051	.000	.004	.001	.000	.006	.016	.000	.026	.502	.000	.001	.001	.000	.005	.141	.000	.294
		100	.409	.000	.009	.001	.001	.063	.219	.000	.110	.768	.000	.001	.001	.000	.135	.398	.000	.234
		40	.708	.023	.071	.006	.002	.064	.433	.000	.101	.888	.027	.047	.003	.001	.092	.493	.000	.145
Moderate (.70..90)	.25	400	.078	.000	.000	.001	.000	.003	.068	.000	.006	.248	.000	.000	.000	.000	.033	.016	.000	.201
		100	.356	.000	.000	.003	.000	.022	.271	.000	.036	.610	.000	.000	.002	.000	.016	.222	.000	.272
		40	.598	.000	.004	.002	.002	.047	.376	.022	.127	.850	.000	.002	.004	.000	.113	.516	.001	.214
	.20	400	.005	.000	.000	.000	.000	.001	.002	.000	.002	.180	.000	.000	.000	.000	.025	.006	.000	.149
		100	.158	.000	.001	.001	.000	.011	.068	.001	.056	.368	.000	.000	.001	.000	.026	.107	.000	.288
		40	.447	.000	.000	.001	.002	.075	.239	.015	.127	.895	.000	.003	.003	.000	.141	.412	.000	.246
	.05	400	.295	.000	.170	.000	.035	.000	.000	.000	.000	.061	.000	.000	.000	.000	.004	.000	.000	.057
		100	.283	.000	.120	.000	.007	.000	.000	.000	.016	.466	.000	.000	.000	.000	.141	.043	.000	.202
		40	.271	.000	.124	.000	.015	.036	.005	.000	.091	.679	.000	.039	.001	.000	.102	.164	.000	.284
Difficult (.40..29)	.25	400	.479	.000	.000	.005	.000	.013	.451	.000	.019	.587	.000	.000	.003	.000	.052	.137	.000	.320
		100	.638	.000	.000	.006	.000	.009	.596	.011	.058	.751	.000	.000	.003	.000	.113	.419	.000	.216
		40	.888	.000	.000	.000	.001	.036	.636	.026	.091	.876	.000	.000	.001	.000	.086	.635	.000	.151
	.20	400	.061	.000	.000	.000	.000	.001	.045	.000	.015	.408	.000	.000	.001	.000	.046	.053	.000	.388
		100	.368	.000	.000	.004	.000	.016	.259	.003	.095	.685	.000	.000	.001	.000	.109	.281	.000	.294
		40	.648	.000	.004	.002	.002	.040	.447	.017	.136	.843	.000	.000	.006	.000	.099	.543	.001	.199
	.05	400	.178	.000	.135	.000	.013	.000	.000	.000	.000	.125	.000	.000	.000	.000	.017	.004	.000	.186
		100	.182	.000	.101	.000	.045	.003	.000	.000	.013	.541	.000	.003	.001	.000	.158	.002	.000	.297
		40	.242	.000	.118	.000	.009	.031	.011	.009	.093	.795	.000	.037	.003	.000	.214	.147	.000	.384

Table II: Proportions related to the various types of errors, detected by using procedure B, over the various combinations of validity, skill difficulty, probability of guessing and sample size levels.

Note: 1. For each dataset, proportion of "No decisions" in the sum of the proportions related to the various types of errors.
2. This table is to be used together with Table III.

The error type "negative variance of $\hat{P}_{II}$" was defined differently in procedure B (as compared with procedure A). In procedure A, this error type was checked, after the estimate of P_{II} was obtained, at the end of the fifth iteration. However, in procedure B, check for convergence was made after each iteration. This error type was checked whenever the convergence of the estimate of P_{II} was obtained. Since ten iterations rather than five were allowed in procedure B, this error type occurred somewhere in between the first and the tenth iteration rather than at the end of the fifth iteration. As shown in Table XI, this error type was detected mainly in the valid datasets. It was explained in section 3.1 (where error types detected in procedure A were discussed), that the more frequent occurrences of this error type ("negative variance of $\hat{P}_{II}$") could be due to the fact that, in the valid datasets, the P_{II} estimate values are more likely to assume small negative values.

In error category II, the two error types detected in procedure B, are "non-convergence of estimates" and "unreasonably large estimates". Both error types are related to the problems of convergence of parameter estimates. The first type was detected at the end of the tenth iteration of the parameter estimation process. The second type was detected during the parameter estimation

process by means of a range check (-1, 2) which was applied to each of the intermediate estimates.

In Table XI, it is shown that the "non-convergence of estimates" error type was detected more frequently over the invalid datasets. The explanations given to "non-convergence" cases in section 3.1 can also be used here to explain this pattern. However, it may be noted that the magnitudes of the proportions due to this error type (non-convergence of estimates) are smaller in the results obtained from using procedure B as compared with the corresponding results obtained from using procedure A. One explanation for this phenomenon is that there were more iterations allowed in procedure B (10 rather than 5). Another explanation is that the range check for "unreasonably large estimates" was utilized in procedure B to detect the cases which would potentially lead to "non-convergence" of parameter estimates. If this range check were removed, the proportions of "non-convergence" of estimates would be increased.

Over the valid datasets, the proportions, related to this error type, range from .000 to .094 . The lower error proportions occurred in datasets associated with sample sizes of 400. The higher error proportions were related to small sample sizes of 40 and easy skill hierarchies. Over

the invalid datasets, the proportions range from .004 to .214 . The higher proportions also tend to occur with small sample sizes of 40 and 100. The lower proportions were associated with sample sizes of 400 and small probability of guessing level of .05 . The above patterns of error proportions were somewhat similar to those found in procedure A, except that, the magnitudes of the proportions of "non-convergence" in the results from using procedure A were of larger magnitudes.

Also, in Table XI, it is shown that the "unreasonably large estimates" error type was generally more often detected over the invalid datasets.

Over the valid datasets, the lower error proportions, for this error type, appear to be associated with sample sizes of 400, low probability of guessing of .05 and with moderate skill learning hierarchies. The high error proportions appear to be due to small sample sizes of 40. High proportions also seem to be related to easy skills hierarchies.

Over the invalid datasets, similar patterns may be observed, except that the proportions are generally higher than the valid counterparts.

In error category III, two error types are identified,

namely, "estimate $\hat{P}_{II}$ being a large negative value" and "other estimates being outside the (0,1) range". Both error types are related to improperly convergent parameter estimates.

The first error type, in category III, was rarely detected in the invalid datasets. It is reasonable to expect that the error proportions of this type of error, were higher in the valid datasets because the population values of P_{II} over the valid datasets are all equal to zero. It may be observed, in Table XI, over the proportions in the valid datasets, that this error type was not detected in large sample sizes of 400 and was never detected in datasets with probability of guessing of .05 .. It is generally expected that with large sample sizes and low measurement error of probability of guessing, the parameter estimation would be more accurate. According to the nature of this error type (detected when $\hat{P}_{II}$ is less than -.05), it is reasonable to find that this error type never occurred in samples sizes of 400 and guessing probability of .05 .

The second error type, in category III, is related to improperly convergent estimates other than $\hat{P}_{II}$ (There were six other parameter estimates). The population parameter values of the other estimates range from .02 to .95 . The check utilized to detect this error type was a very liberal

range check of (-.05 to 1.05). Since the results on this error type were related to 6 parameter estimates, it is difficult to provide clear cut explanations on the observed patterns of proportions. More data may be useful for better understanding of the patterns. However, it may be noted that this error type was detected more frequently in the invalid datasets. Over the invalid datasets, the proportions range from .057 to .320 and over the valid datasets, the proportions range from .000 to .136 .

So far, only the first part of the results obtained from using procedure B has been presented and discussed. The second part of the results are related to correct and incorrect decisions at the various decision criteria. Brief discussions are given in the next paragraph.

As shown in Table XII, over the valid datasets, the proportions of correct decisions decrease as the width of the confidence interval decreases. However, over the invalid datasets, the proportions of correct decisions increase as the width of the confidence interval decreases. Also, the proportions of incorrect decisions are generally higher over the invalid datasets than over the valid datasets. The above phenomena were discussed when the results obtained from procedure A were studied.

Skill Difficulty Levels (P_I, P_{II})	Probabilities of Guessing, θ	Sample Size, N	Proportions of "No Decisions"	Datasets for VALID hierarchies: 95% Conf. Int. C	95% Conf. Int. I	90% Conf. Int. C	90% Conf. Int. I	85% Conf. Int. C	85% Conf. Int. I	80% Conf. Int. C	80% Conf. Int. I	75% Conf. Int. C	75% Conf. Int. I	70% Conf. Int. C	70% Conf. Int. I	Proportions of "No Decisions"	Datasets for INVALID hierarchies: 95% Conf. Int. C	95% Conf. Int. I	90% Conf. Int. C	90% Conf. Int. I	85% Conf. Int. C	85% Conf. Int. I	80% Conf. Int. C	80% Conf. Int. I	75% Conf. Int. C	75% Conf. Int. I	70% Conf. Int. C	70% Conf. Int. I
		400	.305	.615	.030	.613	.002	.609	.006	.606	.011	.599	.020	.575	.049	.843	.004	.453	.034	.023	.863	.394	.111	.246	.169	.288	.217	.240
	.25	100	.744	.236	.000	.236	.000	.236	.000	.235	.001	.232	.004	.226	.010	.871	.009	.120	.013	.116	.014	.115	.015	.116	.019	.110	.023	.106
		40	.883	.115	.002	.115	.002	.115	.002	.114	.003	.112	.005	.112	.005	.899	.024	.077	.026	.075	.035	.066	.066	.035	.068	.033	.068	.033
Easy		400	.251	.749	.000	.746	.003	.743	.006	.734	.015	.712	.037	.691	.058	.493	.017	.490	.077	.430	.133	.374	.203	.304	.253	.254	.309	.198
(.95..90)	.20	100	.697	.303	.000	.303	.000	.303	.000	.301	.002	.296	.007	.295	.008	.863	.011	.146	.015	.142	.016	.141	.029	.127	.026	.131	.031	.126
		40	.847	.151	.002	.151	.002	.151	.002	.147	.006	.147	.006	.147	.006	.890	.025	.085	.029	.081	.034	.076	.072	.038	.075	.035	.075	.035
		400	.051	.949	.000	.946	.003	.941	.008	.927	.022	.919	.030	.908	.041	.502	.203	.295	.338	.160	.485	.093	.639	.059	.666	.034	.677	.023
	.05	100	.489	.591	.000	.591	.000	.591	.000	.586	.005	.580	.011	.579	.012	.761	.012	.227	.023	.216	.030	.209	.061	.198	.066	.173	.089	.150
		40	.780	.299	.001	.298	.002	.298	.002	.297	.003	.297	.003	.297	.003	.808	.044	.148	.052	.140	.064	.128	.110	.082	.117	.070	.116	.070
		400	.870	.920	.002	.909	.013	.898	.024	.876	.046	.852	.070	.823	.099	.248	.390	.362	.602	.150	.699	.063	.731	.031	.738	.024	.763	.009
	.25	100	.354	.644	.000	.644	.000	.639	.005	.624	.020	.605	.039	.597	.047	.610	.002	.388	.019	.371	.053	.337	.095	.295	.132	.258	.183	.207
		40	.598	.402	.000	.402	.000	.401	.001	.378	.006	.391	.009	.390	.012	.850	.001	.149	.001	.149	.002	.148	.008	.142	.016	.134	.023	.127
Moderate		400	.805	.991	.004	.982	.013	.961	.034	.941	.054	.924	.071	.899	.096	.188	.691	.129	.790	.030	.812	.008	.817	.003	.820	.000	.820	.000
(.70..50)	.20	100	.158	.842	.000	.840	.002	.838	.004	.830	.012	.818	.024	.787	.055	.569	.013	.427	.063	.377	.129	.311	.196	.244	.283	.177	.316	.124
		40	.467	.533	.000	.533	.000	.533	.000	.521	.007	.527	.006	.519	.014	.805	.000	.195	.000	.195	.005	.190	.010	.185	.024	.171	.047	.148
		400	.205	.795	.000	.795	.000	.785	.010	.775	.020	.763	.032	.739	.056	.061	.938	.001	.939	.000	.939	.000	.939	.000	.939	.000	.939	.000
	.05	100	.283	.717	.000	.716	.001	.716	.001	.711	.006	.709	.008	.707	.010	.466	.298	.236	.464	.098	.581	.033	.512	.027	.521	.013	.527	.007
		40	.271	.729	.000	.729	.000	.729	.000	.728	.001	.727	.002	.726	.003	.879	.000	.313	.030	.283	.074	.247	.127	.199	.177	.144	.240	.117
		400	.479	.516	.007	.504	.017	.487	.016	.479	.042	.461	.060	.448	.073	.507	.117	.376	.259	.234	.354	.139	.411	.082	.450	.043	.478	.021
	.25	100	.638	.369	.001	.369	.001	.364	.006	.357	.013	.352	.018	.341	.029	.751	.002	.247	.005	.244	.016	.235	.035	.214	.057	.192	.076	.175
		40	.800	.200	.000	.200	.000	.199	.001	.197	.003	.194	.006	.191	.009	.876	.000	.124	.001	.123	.003	.121	.005	.119	.011	.113	.016	.108
Difficult		400	.061	.913	.006	.921	.018	.901	.038	.882	.057	.859	.080	.838	.101	.408	.334	.258	.476	.116	.538	.054	.564	.026	.583	.009	.584	.008
(.60...25)	.20	100	.369	.632	.000	.630	.002	.623	.009	.613	.019	.600	.032	.575	.057	.685	.061	.312	.022	.293	.044	.271	.091	.234	.128	.195	.154	.161
		40	.648	.352	.000	.352	.000	.351	.001	.349	.003	.343	.009	.336	.016	.843	.000	.157	.000	.157	.003	.154	.007	.150	.013	.144	.028	.129
		400	.178	.822	.000	.821	.001	.810	.012	.798	.024	.791	.031	.767	.055	.125	.875	.000	.875	.000	.875	.000	.875	.000	.875	.000	.875	.000
	.05	100	.162	.838	.000	.838	.000	.838	.000	.831	.007	.824	.014	.824	.014	.541	.268	.212	.377	.082	.426	.033	.448	.017	.451	.004	.458	.005
		40	.262	.738	.000	.714	.000	.738	.000	.716	.007	.716	.002	.736	.002	.745	.011	.234	.030	.207	.071	.174	.110	.135	.145	.100	.167	.091

Table XII: Proportions of correct and incorrect decisions, made by using procedure B at different confidence intervals, over the various combinations of validity, skill difficulty, probability of guessing and sample size levels.

Note: 1. "C" indicates proportion of correct decisions.
2. "I" indicates proportion of incorrect decisions.

In the next section, results obtained from the two modified versions (A and B) of Owston's procedure for validating learning hierarchies will be compared.

3.4 COMPARISONS OF RESULTS OBTAINED FROM PROCEDURES A AND B

There were only two levels of decision criteria used in procedure A. These decision criteria are 95% and 90% confidence intervals. In this section, the proportions of correct decisions obtained in both procedures (A and B) at the 95% and 90% confidence intervals are compared. The comparison criteria of (.80, .95), used in section 3.2, will also be used here. In addition to the proportion of correct decisions, comparisons will be made on the proportion of "no decisions" due to the error types detected in each of the two procedures.

As additional information, the proportions of correct decisions (obtained in procedure B; shown in Table XII) were adjusted by utilizing the error proportions. The adjustment process will be mentioned in this section. The adjusted proportions of correct decisions obtained from each of the two procedures (A and B) were also compared by means of the (.80, .95) criteria, at each of the 95% and 90% confidence intervals.

At the end of this section, the adjusted proportions of "correct decisions", obtained in procedure B, at each of the 85%, 80%, 75% and 70% confidence intervals, will be shown. The incidences of meeting the (.80, .95) criteria will be briefly discussed.

As shown in Table XIII, the proportions of "no decisions" over the 27 valid and 27 invalid datasets, are generally lower in the results obtained from using procedure B. This implies that, by using procedure B, there were generally fewer incidences of "no decisions" cases. One main reason for this phenomenon is that ten rather than five iterations were allowed in procedure B. The proportion of "non-convergence", for each dataset, was thus reduced in procedure B. Consequently, the proportion of "no decisions" was also smaller.

Also, in Table XIII, it may be observed that at each of the 95% and 90% confidence intervals, the proportions of correct decisions are relatively higher in the results obtained from using procedure B. However, it is indicated in the data in Table XIII that the (.80, .95) criteria were not satisfied.

In Table XIV, comparisons between procedure A and B are made by utilizing the adjusted results. The adjusted

Skill difficulty Levels (D_I, D_{II})	Probability of guessing θ	Sample Size N	Proportions of "No decisions"				Proportions of Correct Decisions							
							95% Confidence Interval				90% Confidence Interval			
			Procedure A		Procedure B		Procedure A		Procedure B		Procedure A		Procedure B	
			Valid Ds	Invalid Ds	Valid Ds	Invalid Ds	Valid Ds	Invalid Ds	Valid Ds	Invalid Ds	Valid Ds	Invalid Ds	Valid Ds	Invalid Ds
Easy (.95,.90)	.25	400	0.794	0.752	0.385	0.543	0.206	0.001	0.615	0.004	0.205	0.028	0.413	0.034
		100	0.911	0.929	0.764	0.871	0.089	0.000	0.236	0.009	0.089	0.000	0.236	0.013
		40	0.948	0.970	0.883	0.899	0.051	0.006	0.115	0.024	0.051	0.009	0.115	0.026
	.20	400	0.694	0.702	0.251	0.493	0.306	0.021	0.749	0.017	0.304	0.066	0.746	0.077
		100	0.866	0.910	0.697	0.843	0.134	0.001	0.303	0.011	0.134	0.001	0.303	0.015
		40	0.918	0.954	0.847	0.890	0.080	0.008	0.151	0.025	0.080	0.013	0.151	0.029
	.05	400	0.215	0.558	0.051	0.502	0.785	0.217	0.949	0.203	0.782	0.326	0.946	0.330
		100	0.605	0.816	0.409	0.761	0.395	0.006	0.591	0.012	0.395	0.017	0.591	0.023
		40	0.784	0.903	0.700	0.808	0.215	0.018	0.299	0.048	0.214	0.026	0.298	0.052
Moderate (.70,.50)	.25	400	0.308	0.330	0.078	0.248	0.690	0.309	0.920	0.390	0.682	0.482	0.909	0.602
		100	0.579	0.688	0.356	0.610	0.421	0.003	0.644	0.002	0.421	0.013	0.644	0.019
		40	0.731	0.853	0.598	0.850	0.269	0.000	0.402	0.001	0.269	0.000	0.402	0.001
	.20	400	0.043	0.279	0.005	0.180	0.953	0.572	0.991	0.491	0.944	0.663	0.902	0.790
		100	0.351	0.617	0.158	0.560	0.649	0.013	0.842	0.013	0.647	0.048	0.840	0.063
		40	0.631	0.811	0.467	0.805	0.369	0.001	0.533	0.000	0.369	0.002	0.533	0.000
	.05	400	0.170	0.141	0.205	0.081	0.830	0.859	0.795	0.938	0.830	0.859	0.795	0.939
		100	0.163	0.402	0.203	0.466	0.837	0.257	0.797	0.298	0.836	0.398	0.796	0.444
		40	0.317	0.697	0.271	0.679	0.683	0.006	0.729	0.008	0.683	0.029	0.729	0.030
Difficult (.40,.75)	.25	400	0.002	0.500	0.479	0.507	0.195	0.093	0.514	0.117	0.190	0.195	0.504	0.259
		100	0.806	0.820	0.630	0.751	0.194	0.000	0.369	0.002	0.194	0.006	0.369	0.005
		40	0.857	0.923	0.800	0.876	0.143	0.000	0.200	0.000	0.143	0.000	0.200	0.001
	.20	400	0.223	0.484	0.061	0.408	0.777	0.282	0.933	0.314	0.761	0.464	0.921	0.474
		100	0.552	0.740	0.368	0.685	0.448	0.004	0.632	0.003	0.446	0.018	0.630	0.022
		40	0.765	0.884	0.648	0.843	0.235	0.000	0.352	0.000	0.234	0.002	0.352	0.000
	.05	400	0.135	0.239	0.178	0.125	0.865	0.761	0.822	0.875	0.864	0.761	0.821	0.875
		100	0.137	0.504	0.162	0.541	0.863	0.225	0.838	0.247	0.863	0.336	0.838	0.377
		40	0.329	0.737	0.262	0.755	0.671	0.011	0.738	0.011	0.671	0.047	0.738	0.038

Table IIII: Comparisons of the proportions of "No decisions" and the proportions of correct decisions at 95% and 90% confidence interval, for procedures A and B, over the various combinations of validity, skill difficulty, probability of guessing and sample size levels.

Note: "Ds" indicates datasets; for example "Valid Ds" indicates datasets that characterize valid hierarchies.

Skill Difficulty Levels (D_I, D_{II})	Probability of Guessing: θ	Sample Size: N	Adjusted Proportions of "Correct Decisions"							
			95% Confidence Interval				90% Confidence Interval			
			procedure A		procedure B		procedure A		procedure B	
			valid D_s	invalid D_s	valid D_s	invalid D_s	valid D_s	invalid D_s	valid D_s	invalid D_s
Easy (.95,.90)	.25	400	0.206	0.753	0.615	0.547	0.205	0.780	0.613	0.577
		100	0.097	0.923	0.230	0.874	0.097	0.923	0.230	0.870
		40	0.079	0.946	0.145	0.892	0.079	0.949	0.145	0.894
	.20	400	0.306	0.772	0.749	0.510	0.304	0.767	0.746	0.570
		100	0.138	0.904	0.307	0.848	0.138	0.904	0.307	0.852
		40	0.115	0.934	0.187	0.806	0.115	0.939	0.187	0.890
	.05	400	0.709	0.774	0.953	0.704	0.786	0.883	0.950	0.839
		100	0.400	0.820	0.601	0.772	0.400	0.831	0.601	0.705
		40	0.293	0.871	0.372	0.804	0.292	0.879	0.371	0.813
Moderate (.70,.50)	.25	400	0.698	0.647	0.920	0.638	0.682	0.820	0.909	0.650
		100	0.421	0.691	0.648	0.612	0.421	0.701	0.648	0.629
		40	0.224	0.851	0.430	0.848	0.224	0.851	0.430	0.848
	.20	400	0.953	0.851	0.991	0.871	0.944	0.942	0.982 *	0.920
		100	0.650	0.630	0.844	0.573	0.648	0.665	0.842	0.625
		40	0.379	0.807	0.558	0.802	0.379	0.810	0.558	0.803
	.05	400	1.000 *	1.000	1.000 *	0.999	1.000 *	1.000	1.000 *	1.000
		100	0.957	0.739	0.984	0.764	0.956	0.800	0.983	0.910
		40	0.810	0.644	0.868	0.648	0.810	0.607	0.860	0.670
Difficult (.40,.25)	.25	400	0.195	0.673	0.514	0.624	0.190	0.775	0.504	0.766
		100	0.194	0.820	0.380	0.753	0.194	0.826	0.380	0.756
		40	0.143	0.923	0.229	0.876	0.143	0.923	0.229	0.877
	.20	400	0.772	0.766	0.933	0.742	0.761	0.808	0.921	0.884
		100	0.448	0.744	0.635	0.688	0.446	0.758	0.633	0.707
		40	0.240	0.884	0.375	0.842	0.239	0.886	0.375	0.842
	.05	400	1.000 *	1.000	1.000 *	1.000	0.999 *	1.000	0.999 *	1.000
		100	0.965	0.806	0.984	0.785	0.963	0.917	0.984	0.915
		40	0.792	0.711	0.865	0.729	0.792	0.747	0.865	0.756

Table III: Comparisons of the adjusted (to include some of the "No decisions" cases as "correct decisions" cases) proportions of "correct decisions" at 95% & 90% confidence intervals, for procedures A and B, over the different combinations of validity, skill difficulty, probability of guessing and sample size levels.

Note: "D_s" indicates "datasets"

* indicates that the two criteria, requiring the proportions of "correct decisions" to be larger that .80 for valid hierarchies and to be larger than .75 for invalid hierarchies, were simultaneously satisfied.

results related to procedure A were retrieved from Table X. The adjusted results from procedure B were obtained by the process described in the next paragraph.

Recall that in discussing the various types of errors detected by using procedure B, it was suggested (directly or indirectly) that some error types mainly occurred over the valid datasets while the rest occured chiefly over the invalid datasets. The three error types, namely, "probability values tend to be zero", "negative variance of $\hat{P}_{II}$", and "convergent P_{II} estimate being a large negative number", were more often detected over the valid datasets. Each of these error types was treated as an indication of a valid learning hierarchy. For each valid dataset, the proportion of correct decisions at each of the decision criteria was adjusted upward to include the proportions due to these three error types. There was no change in the proportions of "correct decisions" over the invalid datasets. The other five error types, namely, "marginal frequency equals zero",... etc., were detected more often over the invalid datasets. Each of these five error types was considered globally as an indication of an invalid hierarchy. For each invalid dataset, the proportion of correct decisions at each of the decision criteria was adjusted upward to include the proportions due to these five error types. The adjusted proportions of "correct

decisions" in procedure B were shown in Table XIV and Table XV.

It may be observed, in Table XIV, that at 95% confidence interval, with the adjusted results the (.80, .95) criteria were satisfied twice in the results obtained from each of the two procedures. However, at the 90% confidence interval, these criteria were satisfied three times in the adjusted results obtained from procedure B(as can be observed in the last two columns of Table XIV).

In Table XV, it is shown that, as the width of the confidence interval decreases, the (.80, .95) criteria tended to be more often met. At each of the 75% and 70% confidence intervals, in the adjusted results obtained from procedure B, the (.80, .95) criteria were satisfied eight times. Of these eight times, six occurred over sample sizes of 400 and two occurred over sample sizes of 100. With respect to measurement error of guessing, of these eight times, five occurred over the low guessing level of .05, two over the level of .20 and one over the level of .25 . As far as skill difficulty level is concerned, of these 8 times, four occurred in datasets associated with moderate skills learning hierarchies, three in difficult skills learning hierarchies and one in the easy skills learning hierarchies.

Skill Difficulty Levels (D_I, D_{II}) P	Probability of guessing, θ	Sample Size, N	Adjusted Proportions of "Correct Decisions" for procedure B							
			85% Confidence Interval		80% Confidence Interval		75% Confidence Interval		70% Confidence Interval	
			Valid Datasets	Invalid Datasets	Valid Datasets	Invalid Datasets	Valid Datasets	Invalid Datasets	Valid Datasets	Invalid Datasets
Easy (.95,.90)	.25	400	0.609	0.606	0.604	0.654	0.595	0.712	0.575	0.760
		100	0.238	0.879	0.237	0.880	0.234	0.884	0.228	0.888
		40	0.143	0.903	0.144	0.934	0.142	0.936	0.142	0.936
	.20	400	0.743	0.676	0.734	0.696	0.712	0.746	0.691	0.802
		100	0.307	0.853	0.305	0.857	0.300	0.863	0.299	0.868
		40	0.187	0.895	0.183	0.933	0.183	0.936	0.183	0.936
	.05	400	0.945	0.904	0.931	0.940	0.923 *	0.965	0.912 *	0.970
		100	0.601	0.790	0.596	0.801	0.590	0.826	0.589	0.849
		40	0.371	0.826	0.370	0.870	0.370	0.872	0.370	0.874
Moderate (.70,.50)	.25	400	0.894	0.917	0.876 *	0.979	0.852 *	0.986	0.823 *	0.991
		100	0.643	0.663	0.628	0.695	0.609	0.742	0.601	0.793
		40	0.429	0.849	0.426	0.893	0.421	0.863	0.418	0.870
	.20	400	0.961 *	0.992	0.941 *	0.997	0.924 *	1.000	0.899 *	1.000
		100	0.840	0.609	0.832	0.756	0.812	0.823	0.789	0.874
		40	0.558	0.807	0.536	0.812	0.532	0.826	0.544	0.849
	.05	400	0.998 *	1.000	0.988 *	1.000	0.988 *	1.000	0.984 *	1.000
		100	0.983 *	0.967	0.978 *	0.978	0.976 *	0.987	0.974 *	0.993
		40	0.868	0.714	0.867	0.762	0.866	0.817	0.865	0.869
Difficult (.40,.25)	.25	400	0.407	0.861	0.479	0.910	0.461	0.957	0.448	0.977
		100	0.375	0.765	0.368	0.786	0.363	0.808	0.352	0.825
		40	0.228	0.879	0.226	0.881	0.223	0.887	0.220	0.892
	.20	400	0.901	0.946	0.882	0.974	0.859 *	0.991	0.838 *	0.992
		100	0.626	0.729	0.616	0.766	0.603	0.805	0.578	0.839
		40	0.374	0.845	0.372	0.849	0.366	0.855	0.359	0.870
	.05	400	0.988 *	1.000	0.976 *	1.000	0.959 *	1.000	0.945 *	1.000
		100	0.984 *	0.964	0.977 *	0.978	0.970 *	0.989	0.970 *	0.992
		40	0.865	0.789	0.863	0.828	0.863	0.843	0.863	0.809

Table IV: Adjusted proportions of "correct decisions", at 85%,80%,75%,70% confidence intervals, for procedure B, over the different combinations of validity, skill difficulty, probability of guessing and sample size levels.

* indicates that the two criteria, requiring the proportions of "correct decisions" to be larger than .80 for valid hierarchies and to be larger than .95 for invalid hierarchies, were simultaneously satisfied.

Recall that, using the results shown in Table X, Owston claimed that the (.80, .95) criteria were satisfied eight times, at each of the 95% and 90% confidence intervals. By comparing Table X with Table XV, it may be observed that the (.80, .95) criteria were met in the same eight pairs of datasets (one pair contains one valid and the corresponding invalid datasets). This implies that, based on the (.80, .95) criteria, the adjusted results obtained from using procedure B at each of the 75% and 70% confidence intervals, were as good as the results obtained by Owston at each of the 95% and 90% confidence intervals.

So far, the results obtained in this study together with the pertinent data obtained by Owston have been presented and discussed. In the next few pages, a brief summary and some general conclusions are made.

SUMMARY AND CONCLUSIONS

A learning hierarchy is a network of intellectual skills arranged in such a manner that the lower-order skills are hypothesized to be the prerequisites of the higher-order skills. White & Clark proposed a theoretical model to describe a two-skill learning hierarchy. In this model, two questions per skill were used to test skill possession; also, measurement errors were considered.

Owston developed a technique, called MAX, for validating two-skill learning hierarchies. The MAX technique is essentially a statistical procedure for estimating one of the parameters, P_{II}, of the White & Clark Model. This parameter is zero for valid hierarchies and larger than zero (but less than 1) for invalid hierarchies.

An iterative maximum likelihood method of estimation was utilized in the MAX technique. From an observed 3x3 pattern of subjects' responses to two questions per skill, a first approximation of the maximum likelihood estimates was obtained via the marginal frequency approach devised by White & Clark. The value of the P_{II} estimate, contained in the final set of estimates, was used for making a decision on the validity of the related hierarchy. To do this, a

confidence interval was constructed around the final P_{II} estimate. If this confidence interval included a zero value, it was concluded that the hierarchy is valid; otherwise it was concluded that the hierarchy is invalid (note that this conclusion or decision may be correct or incorrect depending on the actual validity of the hierarchy).

Owston then implemented the MAX technique and a data generating procedure in a computer program to test the performance of the MAX technique.

For generating simulated data, Owston specified 48 sets of population parameters. These 48 sets were formed by combining the various levels of validity of hierarchy, skill difficulty and measurement error. From each of these parameter sets, 1000 random samples were drawn at each of four levels of sample sizes of 25, 40, 100 and 400.

Next, Owston set up criteria for evaluating the usefulness of the MAX technique. He required that the proportion of correct decisions should be higher than .80 for 1000 samples drawn from valid hierarchies and .95 for samples drawn from invalid hierarchies. These criteria, (.80, .95), were applied to analyze the usefulness of the MAX technique over the valid and invalid hierarchies associated with different levels of skill difficulty,

measurement error of guessing and sample size.

However, in Owston's computer program, various errors were identified in the part of the program where the MAX technique was implemented. These identified errors were grouped into three categories.

Category I error types were associated with problems that caused computer interruptions. Category II error types were related to the problems of non-convergence of parameter estimates in the iterative parameter estimation process. Category III error types were related to improperly convergent estimates.

The identification of the various error types leads the researcher to question the results obtained by Owston.

Two questions about the MAX technique were then raised. The first question was about the usefulness of the MAX technique when the identified errors were considered. The second question was about the possibilities of improving the implementation procedure of the MAX technique.

Corresponding to these two questions, two procedures, A and B, were developed. Procedure A was identical to Owston's procedure for implementing the MAX technique except that error checks were installed for counting the frequency of occurrence of the various identified error types. Two levels of decision criteria, the 95% and 90% confidence

intervals, were used in procedure A. Procedure B was designed to improve upon procedure A; ten rather than five iterations were used in procedure B; also four extra levels of decision criteria, the 85%, 80%, 75% and 70% confidence intervals were used in procedure B.

The design for collecting data, in this study, was similar to Owston's design. From Owston's 48 sets of population parameters, 18 sets were chosen. These 18 sets were formed by combining three levels of skill difficulty and three levels of measurement error of guessing, at each of the two levels of validity of learning hierarchy(valid and invalid). The three levels of skill difficulty were easy(0.95, 0.90), moderate(0.70, 0.50) and difficult(0.40, 0.25); the three levels of measurement errors(guessing probabilities) were .05, .20 and .25 .

From four levels of sample sizes used by Owston, three levels, namely, 40, 100 and 400, were selected. 1000 random samples were generated from each set of parameters at each level of sample size. Each 1000 random sample thus generated was called a dataset. Hence, by combining 18 sets of population parameters with three levels of sample sizes, 27 datasets related to valid hierarchies(called valid datasets) and 27 corresponding datasets related to invalid learning hierarchies(called invalid datasets) were formed.

Owston's criteria, (.80, .95), for judging the usefulness of the MAX technique, were also utilized for each pair of datasets(one valid and one corresponding invalid datasets).

One unique feature in the current design is that besides calculating the proportions of correct and incorrect decisions when testing the MAX technique, the proportions due to each type of identified error were also calculated. The sum of the error proportions was called the proportion of "no decisions".

Results obtained from using procedure A were then presented. It was found that the proportions of "no decisions" for each pair of datasets were generally high. By analysing the proportions of "no decisions" according to the error types in the error categories, it was found that error types related to non-convergence of parameter estimates were generally detected more often than other error types. Furthermore, it was observed that non-convergence of estimates occurred more frequently in the invalid datasets.

The (.80, .95) criteria on the proportions of correct decisions were not met in any pair of the 27 pairs of datasets. Based on his results, Owston found that the (.80, .95) criteria were met eight times over these dataset pairs.

In analysing the error proportions, it was identified that some error types occurred more often over the valid datasets, while some error types occurred more often over the invalid datasets. These differences among the error types were utilized to aid in validity decision making. An extra set of results was obtained via an adjusting process, in which each error type was either classified as an indication of a valid hierarchy or an invalid hierarchy.

The adjusted results obtained from using procedure A were also compared with Owston's results. It was found that, with the adjusted results, the (.80, .95) criteria were met in two pairs of datasets at each of the 95% and 90% confidence intervals.

Since Owston's results were different from results obtained from using procedure A, before and after adjustments, it was concluded that Owston's results, over the 27 pairs of datasets used in this study, were questionable.

Then, the results obtained from using procedure B was analysed. The proportions of "no decisions" (due to various error types), detected in procedure B, over each pair of datasets, were compared with the corresponding proportions obtained in procedure A. It was found that lower proportions of "no decisions" were detected in procedure B.

One explanation for this phenomenon was that more iterations were allowed in procedure B.

At each of the 95% and 90% confidence intervals, the proportions of correct decisions, obtained in procedures A and B, were also compared. It was noted that, generally, over each dataset, the proportion of correct decisions obtained in procedure B was higher than those obtained in procedure A. This indicates that procedure B was in fact an improved version of the procedure for implementing the MAX technique.

With the error proportions being utilized, the results obtained from procedure B were adjusted. These adjusted results were compared with the corresponding adjusted results obtained from procedure A. At the 90% confidence interval, the (.80, .95) criteria were met, in procedure B results, in three pairs of datasets.

The adjusted results obtained from procedure B, at 85%, 80%, 75% and 70% confidence intervals were also presented and discussed. The (.80, .95) criteria tended to be more often met at the narrower confidence intervals. At each of the 75% and 70% confidence intervals, the (.80, .95) criteria were met eight times over 8 pairs of datasets. These 8 pairs of datasets were the same ones in which Owston found that the (.80, .95) criteria were met at each of the 95% and 90% confidence intervals.

It was observed that, of these 8 pairs of datasets, 6 pairs were associated with sample sizes of 400, 2 pairs were associated with sizes of 100 . Based on these observations, it is recommended that in utilizing the MAX technique for validating learning hierarchies, large sample sizes should be used; preferably, the sizes of 400. Generally sample sizes of 40 or less should not be used.

Besides sample size, skill difficulty and measurement error of guessing are two factors that also influence the performance of the MAX technique. It was found that, at 75% and 70% confidence intervals, the (.80, .95) criteria were met in four dataset pairs associated with moderate skill hierarchies, in three dataset pairs associated with difficult skill hierarchies, and only in one dataset pair associated with easy skill learning hierarchies. For easy skills, such as those found in mastery tests, the MAX technique is generally not recommended, except when the probability of guessing is about .05 and when large sample sizes are used.

With respect to the measurement error of guessing, it was noted that at each of the 75% and 70% confidence intervals, the (.80, .95) criteria were satisfied five times when the guessing probabilities were very small; the (.80, .95) criteria were satisfied twice in dataset pairs

associated with measurement error of guessing of .20, and only once with guessing probability of .25 . This implies that the use of the technique is questionable when using multiple choice test questions with five or four options.

Overall speaking, MAX can be a useful technique for validating learning hierarchies if narrower confidence intervals(such as 75% or 70%) were used, and if errors identified in the technique were utilized to aid in the decision making process.

As has been discussed, non-convergence of parameter estimates is one of the most frequently detected error type in the MAX technique. It is suggested that future researchers may find it interesting to resolve the problems associated with non-convergence of estimates in the MAX technique. Perhaps, if some other methods were used to obtain an initial set of parameter estimates from a 3x3 pattern of subject responses, the frequency of occurrence of the various error types might be drastically reduced. If that were the case, then the proportions of correct decisions and consequently the performance of the MAX technique might be improved.

A probabilistic technique for validating learning hierarchies

BIBLIOGRAPHY

Allendoerfer, C.B., & Oakley, C.O. Principles of Mathematics (2nd ed.). McGraw-Hill Book Company, Inc., 1963.

Cotton, J.W., Gallaghar, J.P., & Marshall, S.P. The Identification and Decomposition of Hierarchical Tasks. American Educational Research Journal, Summer 1977, 14 (3), 189-212.

Durell, A.B. A computer Simulation Study of Measures for Validating Learning Hierarchies. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, April 1974.

Elandt-Johnson, R.C. Probability Models and Statistical Methods In Statistics. New York: Wiley, 1971.

Gagne, R.M. The acquisition of knowledge. Psychological Review, 1962, 69 (4), 355-365.

Gagne, R.M. The Conditions of Learning. New York: Holt, Rinehart, & Winston, 1965.

Gagne, R.M. Learning Hierarchies. Educational Psychologist, November 1968, 6 (1), 1-9.

Gagne, R.M., & White, R.T. Past and Future Research on Learning Hierarchies. Educational Psychologist, 1974, 11 (1), 19-28.

Keith, V. Design and Analysis in Experimentation. Ottawa: University of Ottawa Press, 1972.

Miller, I., & Freund, J.E. Probability and Statistics for Engineers. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.

Owston, R.D. A Comparison of Learning Hierarchy Validation Techniques using Simulated Data. Unpublished doctoral dissertation, University of Ottawa, 1978.

Rao, C.R. Linear Statistical Inference and its Applications. New York: Wiley, 1968.

White, R.T., & Clark, R.M. A test of inclusion which allows for errors of measurement. Psychometrika, March 1973, 38 (1), 77-86.

White, R.T. Indexes Used in Testing the Validity of Learning Hierarchies. J. of Research in Science Teaching, 1974, 11 (1), 61-66.

APPENDICES

Appendix A

Determination of population parameters from difficulty levels of skills I and II

The difficulty of skill I, D_I, and the difficulty of skill II, D_{II}, can be expressed in terms of the four population parameters, P_O, P_I, P_{II} and P_B. P_O is the proportion of subjects possessing neither skill I nor skill II, P_I is the proportion possessing only skill I, P_{II} is the proportion possessing only skill II, P_B is the proportion possessing both skill I and skill II.

$D_I = P_I + P_B$

= proportion possessing skill I,

$D_{II} = P_B + P_{II}$

= proportion possessing skill II.

$\bar{D}_I = 1 - D_I$, and

$\bar{D}_{II} = 1 - D_{II}$.

Beginning with a choice of D_I, for example, to be 0.95, D_{II} can be determined via an assumption that in a valid hierarchy, the phi correlation coefficient, ϕ, is approximately 0.68, and that $P_{II}=0$. Solving a formula $\phi^2 = \frac{\bar{D}_I D_{II}}{\bar{D}_{II} D_I}$ for D_{II}, $D_{II} = 0.898 \doteq 0.90$.

The four P's can hence be found using the following formulae:

$P_O = \bar{D}_I = 0.05$,

$P_I = D_I - D_{II} = 0.05$,

$P_{II} = 0$,

$P_B = D_{II} = 0.90$.

Appendix A

The P's for the corresponding invalid population (e.g. with $\phi = 0.25$) can be obtained by letting $P_{II} = x$, and $\phi = 0.25$, where $x = \bar{D}_I D_{II} - \phi\sqrt{D_I \bar{D}_I D_{II} \bar{D}_{II}}$. Solving for x using the previous D_I, D_{II} values, $x = 0.287 \doteq .03$.

The other P values are found using formulae

$$P_O = \bar{D}_I - x = .02 \quad ,$$

$$P_I = D_I - D_{II} + x = .08 \quad ,$$

$$P_B = D_{II} - x = .87 \quad .$$

At the skill difficulties of (.95, .90), there are four pairs of populations at the four levels of errors of guessing; however the P values for valid and invalid populations have definite patterns as follow*:

	P_O	P_I	P_{II}	P_B
valid	0.05	0.05	0.00	0.90
invalid	0.02	0.08	0.03	0.87

* Same results as obtained by Owston (p.96, set#1 and #9); the formulae derivations are based on Keith, p.67.

Appendix B

Obtaining an initial set of parameter estimates

Owston used the marginal frequencies method to obtain an initial set of parameter estimates required for the process of scoring. He followed the procedures devised by White & Clark. White & Clark set $\hat{P}_{II}=0$, $\hat{\theta}_b=0$ and $\hat{\theta}_c=1$. That is, the proportion possessing skill II without possessing skill I is set to zero, the probability of guessing skill I items is set to zero, and for those who possess skill II, their chance of answering skill II test items correctly are assumed to be 1. Then White & Clark derived formulae to express the other parameter estimates in terms of the marginal frequencies. With an observed 3x3 table as shown in Figure B (same 3x3 table used by Owston for illustrating the MAX technique), the following expressions can be formed. When the marginal frequencies, (a,b,....,f), are substituted into the expressions, numerical values of parameter estimates can be found.

So, $\hat{\theta}_a = \dfrac{2a}{2a+b} = 0.9731$,

$\hat{\theta}_d = \dfrac{e}{e+2f} = 0.1929$,

$\hat{P}_B = 1 - \dfrac{(e+2f)^2}{4fN} = 0.5796$,

$\hat{P}_O = 1-(P_B+P_I) = 0.0886$.

Appendix B

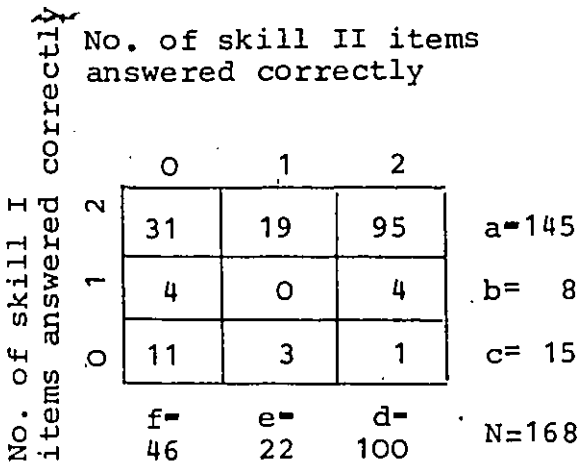
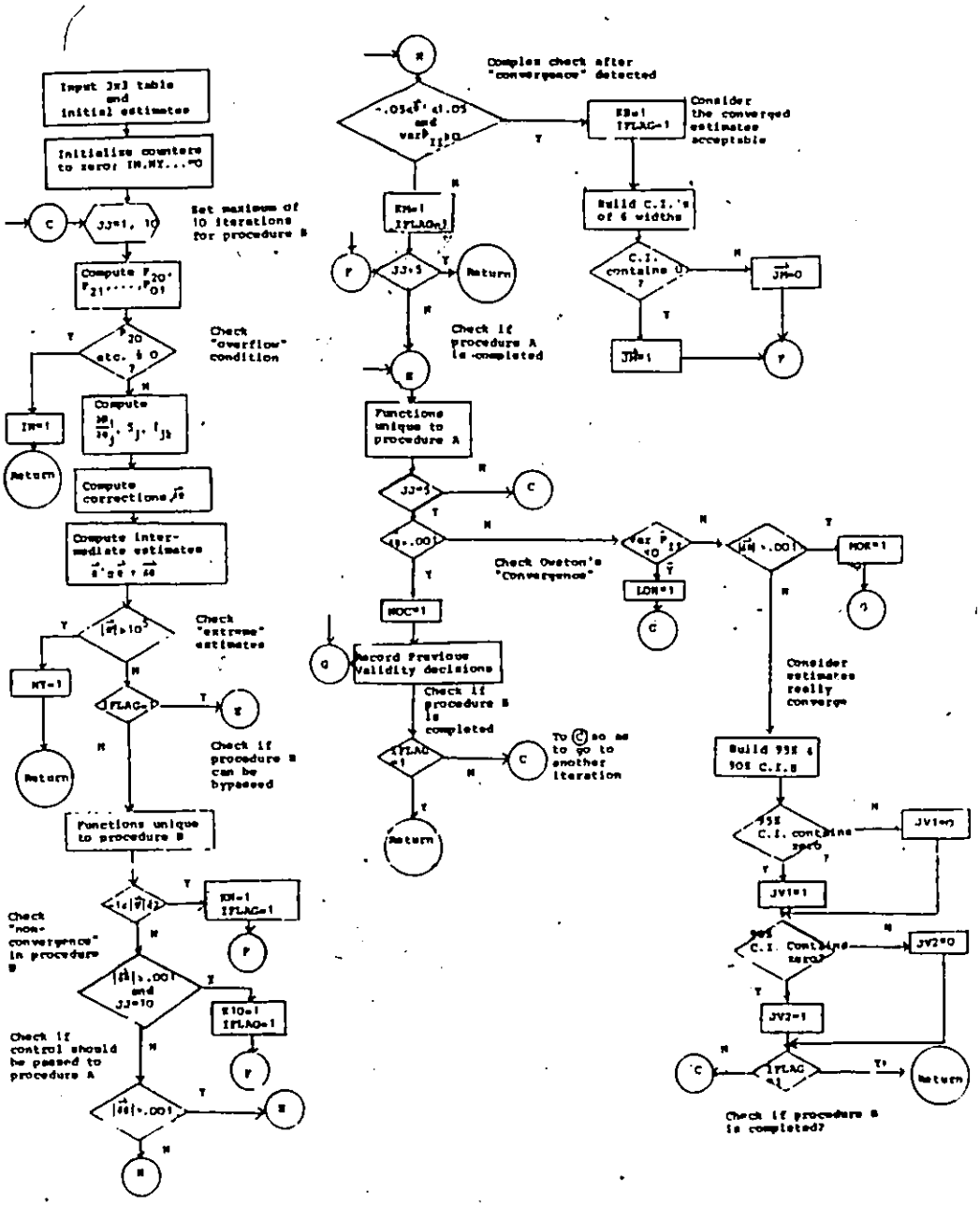


Figure 3: Hypothetical observed frequencies used in illustrating the method of calculating initial parameter estimates

Appendix C

Flowchart to show the various checks implemented in procedure A & B



Appendix D

```
C**   MAIN PROGRAM TO TEST THE PERFORMANCE OF MAX TECHNIQUE
C**           REQUIRES SUBROUTINES  GEN, MC, MAX.
C******
      DIMENSION M(4),PP0(2),PP1(2),PP2(2),PPB(2),Z(6),JM(6),NJ(6),AST(2)
C**   M FOR SAMPLE SIZES, PP'S FOR CONSTANT VALUES OF P-PARAMETERS,
C**   ONE SET FOR VALID HIERARCHIES,ANOTHER SET FOR INVALID HIERARCHIES;
C**   Z FOR DIFFERENT WIDTHS OF CONFIDENCE INTERVAL;
C**   JM FOR RETURNED VALUES OF DECISION FROM MAX SUBROUTINE;
C**   NJ AS ACCULMULATING COUNTERS;AST FOR STORING ASTERISK AND BLANK.
C**   IND IS AN INDICATOR OF MEETING ERROR CRITERIA;MJ AS COUNTER FOR
C**   INVALID HIERARCHIES; ZN FOR THE NORMAL DEVIATE VALUES
      DIMENSION IND(6),MJ(6),ZN(6),ERR(7)
      DIMENSION IC(10),IE(2),ICC(10),IEE(2),IN(10),KM(12),KN(10)
      DIMENSION INN(10),KMM(12),KNN(10)
C******
C**   NINE OBSERVED FREQUENCY VALUES,CORRESPONDING TO THE 3X3 TABLE,
C**           NAMELY,AF,AE,...CD,TOGETHER WITH SAMPLE SIZES N, ARE
C**           COMMON TO THE MAIN PROGRAM AND SUBROUTINES
C******
      COMMON AF,AE,AD,BF,BE,BD,CF,CE,CD,N
      INTEGER AF,AE,AD,BF,BE,BD,CF,CE,CD,AST,EPP
      DATA AST/'*',' '/
      DATA EPP/'A','B','C','D','E','F','+'/
      REAL*4 MJP(6),NJP(6)
      REAL*8 P0,P1,P2,PB,TA,TB,TC,TD,PP(9),HH(7),PI(8),D
C******
C**   READ:    SEED FOR RANDOM NUMBER GENERATION (IS),HERE IS=692035491
C**            FOUR LEVELS OF SAMPLE SIZE (M), ALSO DENOTED BY (N) IN PGM.
C**            THE POPULATION NUMBER (NP);   DIFFICULTY LEVEL OF SKILL 1 (D1)
C**            DIFFICULTIES OF SKILL 2 (D2); CONSTANT SET OF THETA VALUES,
C**            TTA,TTB,TTC,TTD;  CONSTANT SETS OF P-VALUES FOR VALID AND
C**            INVALID HIERARCHIES,(PP0(1),PP1(1)...,PP0(2),...PPB(2));
C**            SIX DIFFERENT CONFIDENCE INTERVAL WIDTHS; ZN, NORMAL DEVIATES
C******
      READ(5,10)IS,(M(I),I=1,4),NP,D1,D2,TTA,TTB,TTC,TTD,(PP0(I),PP1(I),
     *PP2(I),PPB(I),I=1,2),(Z(I),I=1,6),(ZN(I),I=1,6)
   10 FORMAT(I9,4I3,I2,2F2.2,4F2.2,8F2.2,6F2.2,4F4.2,1X,/,2F4.2)
C******
C**   PRINT HEADING FOR OUTPUT
C******
      WRITE(6,15)NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(9,15)NP,D1,D2,TTA,TTB,TTC,TTD
   15 FORMAT('1','POPULATION NUMBER:  ***   ',I2,3X,'***',//,1X,'DIFFICU
     1LTY LEVEL OF SKILL 1 :',3X,F4.2,/,1X,'DIFFICULTY LEVEL OF SKILL 2
     2:',3X,F4.2,/,1X,'PROBABILITY OF CORRECTNESS:',5X,'TA=',F4.2,4X,'TB
     3=',F4.2,4X,'TC=',F4.2,4X,'TD=',F4.2)
      WRITE(6,25)  (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(9,25)  (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
   25 FORMAT(' ',' PROPORTIONS OF SKILL POSSESSION:',/,5X,'VALID HIER.
     1',14X,'P0=',F4.2,4X,'P1=',F4.2,4X,'P2=',F4.2,4X,'PB=',F4.2,/,5X,'I
     2NVALID HIER.',15X,'P0=',F4.2,4X,'P1=',F4.2,4X,'P2=',F4.2,4X,'PB=',
     3F4.2)
      WRITE(6,152) IS
      WRITE(9,152) IS
  152 FORMAT('0','INITIAL RANDOM NO. IS:',3X,I9)
      WRITE(6,41)M(3)
      WRITE(9,41)M(3)
   41 FORMAT(' ','SAMPLE SIZE IS',6X,'***',3X,I3,2X,'***')
      WRITE(6,42)
      WRITE(9,42)
   42 FORMAT('-',//////////,1X,'SYMBOLS USED TO DENOTE COMPLETION STATUS
     C FOR EACH SAMPLE,',//,4X,'SYMBOL',17X,'DESCRIPTIONS')
      WRITE(6,43)
```

Appendix D

```
      WRITE(9,43)
      WRITE(6,44)
      WRITE(9,44)
      WRITE(6,45)
      WRITE(9,45)
      WRITE(6,46)
      WRITE(9,46)
      WRITE(6,47)
      WRITE(9,47)
      WRITE(6,48)
      WRITE(9,48)
      WRITE(6,49)
      WRITE(9,49)
   43 FORMAT(' ',5X,'A',9X,'MARGINAL COLUMN TOTAL, F ,EQUALS TO ZERO,CAU
     CSING UNDEFINED SITUATION')
   44 FORMAT(' ',5X,'B',9X,'INITIAL PARAMETER ESTIMATE(S) BEING OUTSIDE
     CTHE RANGE (0,1)')
   45 FORMAT(' ',5X,'C',9X,'MAX TECH. INVOLVES EST.(S) P20,P21,...VALUES
     C,ONE OR MORE SUCH VALUE FOUND TO BE ZERO,CAUSING UNDEF. SITUATIONS
     C')
   46 FORMAT(' ',5X,'D',9X,'DURING ITERATIVE PROCESS,EXTREME ESTIMATES O
     CUTSIDE (-1,2) RANGE OCCURRED;SUBSEQUENT CONVG. NOT LIKELY')
   47 FORMAT(' ',5X,'E',9X,'AT END OF 10 ITERATIONS CONVERGENCE WAS NOT
     CATTAINED')
   48 FORMAT(' ',5X,'F',9X,'CONV. WITHIN 10 ITER. BUT ESTIMATES OFF (-0.
     C05,1.05) RANGE OR ERROR VARIANCE FOR P2 <0.')
   49 FORMAT(' ',5X,'*',9X,'CONV. WITHIN 10 ITER. AND MAX TECH. CAN BE M
     CEANINGFULLY APPLIED TO MAKE VALIDITY DECISIONS')
C**    FIRST LOOP TO VARY SAMPLE SIZE AT 4 LEVELS
       N=N(3)
C**    SET CUMULATIVE COUNTERS TO ZERO
C**    FOR EACH SAMPLE SIZE, AFTER GENERATING 1000 SAMPLES ON VALID HIERARCHIES
C**    ANOTHER 1000 SAMPLES WILL BE GENERATED ON INVALID HIERARCHIES
C**    SECOND LOOP FOR VALID AND INVALID HIERARCHIES
       DO 30 J=1,2
       WRITE(6,51)
       WRITE(9,51)
       WRITE(6,36)
       WRITE(9,36)
       WRITE(6,37)
       WRITE(9,37)
       WRITE(6,38)
       WRITE(9,38)
       WRITE(6,39)
       WRITE(9,39)
   51 FORMAT('1','LISTING OF INFORMATION ABOUT EACH SAMPLE GENERATED')
   36 FORMAT('0','1',2X,'EP',1X,'SAM#',1X,'AF',2X,'AE',2X,'AD',2X,'BF',2
     CX,'BE',2X,'BD',2X,'CF',2X,'CE',2X,'CD',2X,'INIT.P0',6X,'P1',6X,'P2
     C',6X,'PB',6X,'TA',6X,'TB',6X,'TC',6X,'TD',11X,'HALT-ITER#')
   37 FORMAT(' ','2',9X,'FINAL P20',10X,'P21',10X,'P22',10X,'P10',10X,'P
     C11',10X,'P12',10X,'P00',10X,'P01',10X,'P02',7X)
   38 FORMAT(' ','3',9X,'FINAL H11(P0)',7X,'H22(P1)',8X,'H33(P2)',8X,'H4
     C4(TA)',8X,'H55(TB)',8X,'H66(TC)',8X,'H77(TD)',6X,'DETERMINANT')
   39 FORMAT(' ','4',9X,'FINAL',2X,'P0',11X,'P1',11X,'P2',11X,'PB',11X,
     C'TA',11X,'TB',11X,'TC',11X,'TD')
       KSS=0
       IEE(1)=0
       IEE(2)=0
       K10K=0
C      CUMULATIVE COUNTERS FOR JC1,JC2,JP1,JP2 AND FREQ. OF PSEUDO CONVERGENCE
       NJC1=0
       NJC2=0
       NJP1=0
       NJP2=0
       MJC1=0
       MJC2=0
```

Appendix D

```
      NJP1=0
      NJP2=0
      NNOK=0
      DO 1 I=1,10
      ICC(I)=0
      IBB(I)=0
      KMM(I)=0
      KNN(I)=0
    1 CONTINUE
      KMM(11)=0
      KMM(12)=0
C**   LOOP FOR INITIALIZING THE COUNTERS TO ZERO AND INDICATOR TO BLANKS
      DO 40 K=1,6
      NJP(K)=0.
      NJP(K)=0.
      NJ(K)=0
   40 NJ(K)=0
C**   THIRD LOOP TO GENERATE 1000 RANDOM SAMPLES OF SIZE N, VALIDITY J
      IND(1)=AST(1)
      IND(2)=AST(2)
      MM=0
      NJ1=0
      NJ2=0
      NOCC=0
      NLOH=0
      NJK1=0
      NJK2=0
      NJH1=0
      NJH2=0
      MJ1=0
      MJ2=0
      MJK1=0
      MJK2=0
      MJH1=0
      MJH2=0
      NNY=0
      NJY1=0
      NJY2=0
      MJY1=0
      MJY2=0
      JV1=0
      JV2=0
C     LOOP TO GENERATE 1000 SAMPLES
      DO 50 K=1,1000
      JJ=0
      NOC=0
      LOH=0
C     JC1,JC2 USED FOR CHECKING PSEUDO CONVERGENCE, ALSO NCK
C     JP1,JP2 USED FOR (P20,P21... CASES)
      JC1=0
      JC2=0
      NOK=0
      JP1=0
      JP2=0
      JY1=0
      JY2=0
      NY=0
C**   RESTORE THE PARAMETER VALUES FOR EACH SAMPLE
      P0=PP0(J)
      P1=PP1(J)
      P2=PP2(J)
      PB=PPB(J)
      TA=TTA
      TB=TTB
      TC=TTC
      TD=TTD
```

Appendix D

```
      IF(K.LE.25) GOTO 17
      GOTO 18
   17 MM=MM+1
      IF(MM.EQ.11)GOTO 82
      GOTO 83
   82 WRITE(6,51)
      WRITE(9,51)
      WRITE(6,36)
      WRITE(9,36)
      WRITE(6,37)
      WRITE(9,37)
      WRITE(6,38)
      WRITE(9,38)
      WRITE(6,39)
      WRITE(9,39)
      MM=1
   18 IF(K.GT.25) GOTO 19
      GOTO 83
   19 MM=MM+1
      IF(MM.EQ.11)GOTO 98
      GOTO 83
   98 WRITE(9,51)
      WRITE(9,36)
      WRITE(9,37)
      WRITE(9,38)
      WRITE(9,39)
      MM=1
C**   CALL GEN SUBROUTINE TO CREATE A 3X3 TABLE OF OBSERVED FREQUENCIES
C**   CALL MC SUBROUTINE TO ARRIVE AT AN INITIAL SET OF STATISTICS TO BE
C**   USED BY THE MAX SUBROUTINE
   83 CALL GEN(P0,P1,P2,PB,TA,TB,TC,TD,IS,IC,IF)
      IF(IE(2).EQ.1)GOTO 2
      PI(1)=P0
      PI(2)=P1
      PI(3)=P2
      PI(4)=PB
      PI(5)=TA
      PI(6)=TB
      PI(7)=TC
      PI(8)=TD
      IF(IC(9).EQ.1)GOTO 3
      GOTO 4
    2 DO 5 I=1,2
      IEE(I)=IEE(I)+IE(I)
    5 CONTINUE
      IF(K.LE.25) GOTO 20
      GOTO 21
   20 WRITE(6,55)IND(J),ERR(1),K,AF,AE,AD,BF,BE,BD,CF,CE,CD,JJ
   55 FORMAT('0','1',1X,A1,1X,A1,I4,9(1X,I3),84X,I2)
   21 WRITE(9,55)IND(J),ERR(1),K,AF,AE,AD,BF,BE,BD,CF,CE,CD,JJ
      GOTO 50
    3 DO 7 I=1,9
      ICC(I)=ICC(I)+IC(I)
    7 CONTINUE
    4 D=0.0
C**   CALL MAX SUBROUTINE,PASSING THE CONFIDENCE INTERVAL WIDTHS AND
C**   RETURN THE CORRESPONDING DECISION VALUES(1 INDICATES VALID, 0 INDICATES
C**   INVALID)
      CALL MAX(P0,P1,P2,PB,TA,TB,TC,TD,ZM,JH,IN,KH,KN,K10,K,JJ,KS,PP,MM,
     CD,J,PI,JV1,JV2,JK1,JK2,JH1,JH2,LOH,NOC,NOK,NY)
      IF(NY.EQ.1)GOTO 75
      GOTO 76
   75 NNY=NNY+NY
      JY1=JV1
      JY2=JV2
      NJY1=NJY1+JY1
```

Appendix D

```
      MJY2=MJY2+JY2
      MJY1=MJY1+1-JY1
      MJY2=MJY2+1-JY2
   76 IF(IN(10).EQ.1)GOTO 8
      IF(KN(9).EQ.1)GOTO 9
      IF(KN(10).EQ.1)GOTO 11
      IF(K10.EQ.1)GOTO 12
      IF(KS.EQ.1) KSS=KSS+KS
      GOTO 62
    8 DO 13 I=1,10
      INN(I)=INN(I)+IN(I)
   13 CONTINUE
      JP1=JV1
      JP2=JV2
      MJP1=MJP1+JP1
      MJP2=MJP2+JP2
      MJP1=MJP1+1-JP1
      MJP2=MJP2+1-JP2
      GOTO 62
    9 DO 14 I=1,9
      KNN(I)=KNN(I)+KN(I)
   14 CONTINUE
      GOTO 62
   11 DO 16 I=1,12
      KMM(I)=KMM(I)+KM(I)
   16 CONTINUE
      GOTO 62
   12 K10K=K10K+K10
C**   DEPENDING ON VALIDITY OF HIERARCHIES, STANDARDS OF CORRECT DECISION ARE
C**   DIFFERENT
   62 MJ1=MJ1+JV1
      MJ2=MJ2+JV2
      MJ1=MJ1+1-JV1
      MJ2=MJ2+1-JV2
      MOCC=MOCC+NOC
      MLOH=MLOH+LOH
      MMOK=MMOK+MOK
      IF(NOC.EQ.1)GOTO 84
      GOTO 85
   84 MJK1=MJK1+JK1
      MJK2=MJK2+JK2
      MJK1=MJK1+1-JK1
      MJK2=MJK2+1-JK2
   85 IF(LOH.EQ.1)GOTO 86
      GOTO 87
   86 MJH1=MJH1+JH1
      MJH2=MJH2+JH2
      MJH1=MJH1+1-JH1
      MJH2=MJH2+1-JH2
   87 IF(MOK.EQ.1)GOTO 91
      GOTO 92
   91 JC1=JV1
      JC2=JV2
      MJC1=MJC1+JC1
      MJC2=MJC2+JC2
      MJC1=MJC1+1-JC1
      MJC2=MJC2+1-JC2
   92 IF(IN(10).EQ.1)GOTO 50
      IF(MY.EQ.1)GOTO 50
      IF(KN(9).EQ.1)GOTO 50
      IF(KN(10).EQ.1)GOTO 50
      IF(K10.EQ.1)GOTO 50
      IF(J.EQ.2) GOTO 90
C**   LOOP TO SUM UP NUMBER OF INCORRECT DECISIONS FOR VALID HIERARCHIES
      DO 60 L=1,6
      MJ(L)=MJ(L)+1-JM(L)
```

Appendix D

```
      MJ(L)=MJ(L)+JM(L)
   60 CONTINUE
      GO TO 50
C**   LOOP TO SUM UP NUMBER OF INCORRECT DECISIONS FOR INVALID HIERARCHIES
   90 DO 80 L=1,6
      MJ(L)=MJ(L)+JM(L)
      MJ(L)=MJ(L)+1-JM(L)
   80 CONTINUE
C**   THIRD LOOP COMPLETED
   50 CONTINUE
      K=K-1
C**   FOR VALID HIERARCHIES
      IF (J.EQ.2) GO TO 100
      NTT=IEE(2)+KNN(9)+KMM(10)+K10K+INN(10)+KNY
      DO 78 L=1,6
      MJP(L)=100*FLOAT(MJ(L))/FLOAT(KSS)
      MJP(L)=100*FLOAT(MJ(L))/FLOAT(KSS)
   78 CONTINUE
      WRITE(6,15) NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(9,15) NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(6,25)  (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(9,25)  (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(6,41) M(3)
      WRITE(9,41) M(3)
      WRITE(6,61)
      WRITE(9,61)
   61 FORMAT('-','SUMMARY OF RESULTS FOR POPULATION CHARACTERISING VALID
     C HIERARCHIES')
      WRITE(6,67) K
      WRITE(9,67) K
   67 FORMAT('0','NO. OF SAMPLES GENERATED',41X,I4,/,1X,'SAMPLES REJECTE
     CD DUE TO ERRORS OF THE FOLLOWING TYPES',/,6X,'ERROR TYPE',5X,'BRIE
     CF DESCRIPTIONS',21X,'FREQ')
      WRITE(6,68) IEE(2),ICC(9),INN(10)
      WRITE(9,68) IEE(2),ICC(9),INN(10)
   68 FORMAT(' ',10X,'A',7X,'F=0',38X,I3,/,11X,'B',7X,'INT. EST. NOT IN
     C(0,1)  (NOT ADDED)',2X,I3,/,11X,'C',7X,'ONE OF P20,P21... EQ.ZERO',
     C16X,I3)
      WRITE(6,69) KNN(9),K10K,KMM(10)
      WRITE(9,69) KNN(9),K10K,KMM(10)
   69 FORMAT(' ',10X,'D',7X,'OFF (-1.0,2.0) RANGE IN ITERATION',8X,I3,/,
     C11X,'E',7X,'NO CONVERGENCE WHEN 10 ITER. COMPLETED',3X,I3,/,11X,
     C'F',7X,'OFF (-0.05,1.05) OR ERR. VAR. FOR P2 <0.',1X,I3)
      WRITE(6,108) KNN(9),KMM(11),KMM(12)
      WRITE(9,108) KNN(9),KMM(11),KMM(12)
  108 FORMAT(' ',18X,'SUPP. INFO. ON TYPE F ERROR',/,19X,'VAR. P2<0.:',4
     CX,I3,/,19X,'P2<-0.05 :',5X,I3,/,19X,'P2>1.05 :',6X,I3)
      WRITE(6,101) KNY
      WRITE(9,101) KNY
  101 FORMAT(' ',10X,'*',7X,'EXTREME PARM.ESTIMATES              ',8X,I3)
      WRITE(6,70) NTT,KSS
      WRITE(9,70) NTT,KSS
   70 FORMAT(' ','TOTAL NO. OF SAMPLES REJECTED',36X,I4,/,/,1X,'NO. OF S
     CAMPLES USED TO INFER POPULATION VALIDITY',17X,I4)
      WRITE(6,71)
      WRITE(9,71)
   71 FORMAT('-','DECISION CRITERIA(CONF. INT.)',1X,5X,'95%',12X,'90%',
     C12X,'85%',12X,'80%',12X,'75%',12X,'70%')
      WRITE(6,72) (MJ(L),MJP(L),L=1,6)
      WRITE(9,72) (MJ(L),MJP(L),L=1,6)
   72 FORMAT(' ','CORRECT DECISIONS(AND %)',6X,6(1X,I4,' (',F6.2,'%)'))
      WRITE(6,73) (MJ(L),MJP(L),L=1,6)
      WRITE(9,73) (MJ(L),MJP(L),L=1,6)
   73 FORMAT(' ','INCORRECT DECISIONS(AND %)',4X,6(1X,I4,' (',F6.2,'%)'
     C)
      WRITE(6,74) (KSS,L=1,6)
```

```
      WRITE(9,74) (KSS,L=1,6)
   74 FORMAT(' ','TOTAL NO. OF DECISIONS(AND %)',1X,6(1X,I4,' (100.00%)'
     C))
      GO TO 88
  100 NTE=IEE(2)+KNN(9)+KMM(10)+K10K+INN(10)+NNY
      DO 79 L=1,6
      NJP(L)=100*FLOAT(NJ(L))/FLOAT(KSS)
      MJP(L)=100*FLOAT(MJ(L))/FLOAT(KSS)
   79 CONTINUE
      WRITE(6,15) NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(9,15) NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(6,25)  (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(9,25)  (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(6,41) M(3)
      WRITE(9,41) M(3)
      WRITE(6,64)
      WRITE(9,64)
   64 FORMAT('-','SUMMARY OF RESULTS FOR POPULATION CHARACTERISING INVAL
     CID HIERARCHIES')
      WRITE(6,67) K
      WRITE(9,67) K
      WRITE(6,68) IEE(2),ICC(9),INN(10)
      WRITE(9,68) IEE(2),ICC(9),INN(10)
      WRITE(6,69) KNN(9),K10K,KMM(10)
      WRITE(9,69) KNN(9),K10K,KMM(10)
      WRITE(6,108) KMM(9),KMM(11),KMM(12)
      WRITE(9,108) KMM(9),KMM(11),KMM(12)
      WRITE(6,101) NNY
      WRITE(9,101) NNY
      WRITE(6,70) NTE,KSS
      WRITE(9,70) NTE,KSS
      WRITE(6,71)
      WRITE(9,71)
      WRITE(6,72) (NJ(L),NJP(L),L=1,6)
      WRITE(9,72) (NJ(L),NJP(L),L=1,6)
      WRITE(6,73) (MJ(L),MJP(L),L=1,6)
      WRITE(9,73) (MJ(L),MJP(L),L=1,6)
      WRITE(6,74) (KSS,L=1,6)
      WRITE(9,74) (KSS,L=1,6)
C**   SECOND LOOP COMPLETED
   88 WRITE(6,15) NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(9,15) NP,D1,D2,TTA,TTB,TTC,TTD
      WRITE(6,25) (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(9,25) (PP0(I),PP1(I),PP2(I),PPB(I),I=1,2)
      WRITE(6,41) M(3)
      WRITE(9,41) M(3)
      IF(J.EQ.2)  WRITE(6,64)
      IF(J.EQ.2)  WRITE(9,64)
      IF(J.EQ.1)  WRITE(6,61)
      IF(J.EQ.1)  WRITE(9,61)
      WRITE(6,77) K
      WRITE(9,77) K
   77 FORMAT('0','NO. OF SAMPLES GENERATED',20X,I4)
      NI1=K-NJ1
      NI2=K-NJ2
      WRITE(6,89) NJ1,NJ2,NI1,NI2
      WRITE(9,89) NJ1,NJ2,NI1,NI2
   89 FORMAT('0','IND. VALID',10X,2(I4,6X),/,1X,'IND. INVAL',10X,2(I4,6X
     C))
      WRITE(6,95) NOCC,NJK1,NJK2,MJK1,MJK2
      WRITE(9,95) NOCC,NJK1,NJK2,MJK1,MJK2
   95 FORMAT('0','NO. OF NON-CONVERGENT CASES',/,10X,I4,/,1X,'IND. VALID
     C',10X,2(I4,6X),/,1X,'IND. INVAL',10X,2(I4,6X))
      WRITE(6,96) NLOH,NJH1,NJH2,MJH1,MJH2
      WRITE(9,96) NLOH,NJH1,NJH2,MJH1,MJH2
   96 FORMAT('0','-VE VARIANCE P2',/,10X,I4,/,1X,'IND. VALID',10X,2(I4,6
```

Appendix D

```
     CX),/,1X,'IND. INVAL',10X,2(I4,6X))
      WRITE(6,63) NMOK,NJC1,NJC2,NJC1,NJC2
      WRITE(9,63) NMOK,NJC1,NJC2,NJC1,NJC2
   63 FORMAT('0','NO. OF PSEUDO CONVERGENT CASES',/,10X,I4,/,1X,'IND. VA
     CLID',10X,2(I4,6X),/,1X,'IND. INVALID',8X,2(I4,6X))
      WRITE(6,66) INN(10),NJP1,NJP2,NJP1,NJP2
      WRITE(9,66) INN(10),NJP1,NJP2,NJP1,NJP2
   66 FORMAT('0','NO. OF CASES P20 OR P21 ETC. =0 ',/,10X,I4,/,1X,'IND.
     CVALID',10X,2(I4,6X),/,1X,'IND. INVALID',8X,2(I4,6X))
      WRITE(6,81) NNY,NJY1,NJY2,NJY1,NJY2
      WRITE(9,81) NNY,NJY1,NJY2,NJY1,NJY2
   81 FORMAT('0','NO. OF EXTREME  PARM. EST. ',/,10X,I4,/,1X,'IND. VALID
     C',10X,2(I4,6X),/,1X,'IND. INVAL',10X,2(I4,6X))
      WRITE(6,65) (IEE(2),L=1,3)
      WRITE(9,65) (IEE(2),L=1,3)
   65 FORMAT('0','NO. OF  P=0 CASES',/,10X,I4,/,1X,'IND. INVALID',8X,2(I
     C4,6X))
      NJ1T=NJ1-NJK1-NJH1-NJC1-NJP1-NJY1
      NJ2T=NJ2-NJK2-NJH2-NJC2-NJP2-NJY2
      MJ1T=NI1-MJK1-MJH1-MJC1-MJP1-IEE(2)-MJY1
      MJ2T=NI2-MJK2-MJH2-MJC2-MJP2-IEE(2)-MJY2
      WRITE(6,97) NJ1T,NJ2T,MJ1T,MJ2T
      WRITE(9,97) NJ1T,NJ2T,MJ1T,MJ2T
   97 FORMAT('0','CAN ONLY CLAIM',2X,'IND. VALID',5X,2(I4,6X),/,1X,'CAN
     CONLY CLAIM',2X,'IND. INVAL',5X,2(I4,6X))
   30 CONTINUE
C**   FIRST LOOP COMPLETED
      WRITE(6,31)
      WRITE(9,31)
   31 FORMAT('1',10(' '))
      STOP
      END
C*****
      SUBROUTINE GEN(P0,P1,P2,PB,TA,TB,TC,TD,IS,IC,IE)
C**   SUBROUTINE GEN  TO GENERATE RANDOM SAMPLES ACCORDING TO SPECIFIED
C*****            POPULATION PARAMETERS
      REAL*8 P0,P1,P2,PB,TA,TB,TC,TD
      COMMON AF,AE,AD,BF,BE,BD,CF,CE,CD,N
      INTEGER AF,AE,AD,BF,BE,BD,CF,CE,CD
C**   ALLOCATE MEMORY SPACE FOR 3X3 TABLE,4QUESTIONS AT 2 SKILL LEVELS,
C**   PROBABILITIES OF CORRECTNESS AND TALLIES OF SCORES OF THE QUESTIONS
      DIMENSION M(3,3),A(4),J(4),IC(10),IE(2)
C**   SET E=0 & F=0 COUNTERS TO ZERO
      IE(1)=0
      IE(2)=0
C**   INITIALIZE COUNTERS ON OFF RANGE P & THETA VALUES
      DO 55 I=1,10
      IC(I)=0
   55 CONTINUE
C**   SET THE 3X3 TABLE INITIALLY WITH ZEROS
      DO 300 I=1,3
      DO 300 K=1,3
      M(K,I)=0
  300 CONTINUE
C**   SET UP CRITERIA FOR COMPARING WITH NUMBER TO BE GENERATED
      RA=P0+P1
      RB=P0+P1+P2
C**   DETERMINE THE TRUE STATE OF SKILL POSSESSION FOR EACH SUBJECT
  305 DO 310 KK=1,N
      CALL RANDU(IS,IS,R)
      IF (R.LE.P0) GO TO 320
      IF (R.LE.RA) GO TO 330
      IF (R.LE.RB) GO TO 340
C**   THIS SUBJECT POSSESSES BOTH SKILLS
C**   HIS CHANCES OF ANSWERING CORRETLY ARE AS FOLLOW
      A(1)=TA
```

Appendix D

```
      A(2)=TA
      A(3)=TC
      A(4)=TC
      GO TO 350
C**   THIS SUBJECT POSSESSES NEITHER OF THE SKILLS
  320 A(1)=TB
      A(2)=TB
      A(3)=TD
      A(4)=TD
      GO TO 350
C**   THIS SUBJECT POSESSSES OF SKILL 1 ONLY
  330 A(1)=TA
      A(2)=TA
      A(3)=TD
      A(4)=TD
      GO TO 350
C**   THIS SUBJECT HAS ONLY SKILL 2
  340 A(1)=TB
      A(2)=TB
      A(3)=TC
      A(4)=TC
C*****
C**   GENERATE RANDOM NO. TO DETERMINE IF QUESTIONS ARE CORRECTLY ANSWERED
C*****
  350 DO 360 I=1,4
      CALL RANDU(IS,IS,R)
C**   WHEN A(I) IS GT OR EQ P ANSWER ASSUMED CORRECT
      IF (R-A(I)) 370,370,380
  370 J(I)=1
      GO TO 360
  380 J(I)=0
  360 CONTINUE
C**   ADD 1 TO THE SUM OF SCORES TO CHANGE ORIGIN FROM 0 TO 1
      K=J(1)+J(2)+1
      L=J(3)+J(4)+1
C**   ENTER THE SCORES INTO THE CELLS OF THE TABLE
      N(K,L)=N(K,L)+1
  310 CONTINUE
C*****
C**   RENAME THE CELLS ACCORDING TO MARGINAL FREQUENCIES SYMBOLS
C*****
      AF=N(3,1)
      AE=N(3,2)
      AD=N(3,3)
      BF=N(2,1)
      BE=N(2,2)
      BD=N(2,3)
      CF=N(1,1)
      CE=N(1,2)
      CD=N(1,3)
C**   MARGINAL FREQUENCIES OF 3X3 CONTINGENCY TABLE FROM SUBROUTINE GEN
      A=AF+AE+AD
      B=BF+BE+BD
      C=CF+CE+CD
      D=AD+BD+CD
      E=AE+BE+CE
      F=AF+BF+CF
C**   CHECK IF MARGINAL COL TOTAL E=0 OR NOT
      IF(E.EQ.0.) IE(1)=1
C**   CHECK IF MARGINAL COLUMN TOTAL F=0 OR NOT
      IF(F.EQ.0.) IE(2)=1
      IF(IE(2).EQ.1)GOTO 50
C**   PARAMETER ASSUMPTIONS OF WC MODEL --P(II),THETA(B),THETA(C)
      P2=0.
      TB=0.
      TC=1.
```

Appendix D

```
C**   THREE PARAMETERS ASSUMED FIVE OTHER ONES CALCULATED AS FOLLOW
      TA=(2*AA)/(2*AA+B)
      TD=E/(E+2*F)
      PB=1.-((E+2*F)**2)/(4*F*N)
      P1=((2*AA+B)**2)/(4*AA*N)-PB
      P0=1.-(P1+PB)
      IF((P0.LT.0.).OR.(P0.GT.1)) IC(1)=1
      IF((P1.LT.0.).OR.(P1.GT.1)) IC(2)=1
      IF((P2.LT.0.).OR.(P2.GT.1)) IC(3)=1
      IF((PB.LT.0.).OR.(PB.GT.1)) IC(4)=1
      IF((TA.LT.0.).OR.(TA.GT.1)) IC(5)=1
      IF((TB.LT.0.).OR.(TB.GT.1)) IC(6)=1
      IF((TC.LT.0.).OR.(TC.GT.1)) IC(7)=1
      IF((TD.LT.0.).OR.(TD.GT.1)) IC(8)=1
      DO 60 I=1,8
      IF(IC(I).EQ.1) IC(9)=1
   60 CONTINUE
   50 RETURN
      END
C     SUBROUTINE MAX
      SUBROUTINE MAX(P0,P1,P2,PB,TA,TB,TC,TD,ZN,JM,IN,KM,KN,K10,K1,JJ,KS
     C,PP,HH,D,J1,PI,JV1,JV2,JK1,JK2,JH1,JH2,LOH,NOC,MCK,NY)
      IMPLICIT REAL*8 (A-Z)
      INTEGER JM,IN,KM,K1,KN,K10,KS,IFLAG,IND,AST(8),NY
      REAL*4 ZN
      INTEGER JV1,JV2,NOC,LOH,JK1,JK2,JH1,JH2,MCK,LY
      DATA AST/'V','I','D','E','F','+','S','C'/
      DOUBLE PRECISION DABS,DSQRT
      DIMENSION S(7),ST(7,1),P(7,1),C(7,1),A(7),PI(8),PP(7)
      DIMENSION H(7,7),L(7),M(7),PP(9),HH(7),PZ2(10),PF(9)
      DIMENSION ZN(6),JM(6),IN(10),KM(12),KN(10),Z(6)
      INTEGER I,J,K,JJ,N,J1
      EQUIVALENCE (S(1),ST(1,1))
      INTEGER AF,AE,AD,BF,BE,BD,CF,CE,CD
      COMMON AF,AE,AD,BF,BE,BD,CF,CE,CD,N
C**   K10 TO INDICATE NO CONVERGENCE AT END OF 10 ITERATIONS
C**   KS=1 TO INDICATE SUCCESSFUL CONVERGENCY
      K10=0
      KS=0
C**   VECTOR  IN  STORES COUNTS OF ERROR OF TYPE  P***=0, IN(10) A FLAG
C**   KN STORES FREQ. OF P & THETA VALUES OFF (-0.5,1.5) RANGE DURING ITERATIONS
C**   KN(9) BEING INDICATOR TO CHECK FOR EXIT FROM ITERATIONS
C**   KM STORES COUNTS OF OFF RANGE (-.5,1.05) FOR P & THETA VALUES, AND
C**   ALSO TEST SS ERROR FOR P2 BEING <0
      DO 5 I=1,10
      PZ2(I)=0.
      IN(I)=0
      KM(I)=0
      KN(I)=0
    5 CONTINUE
      KM(11)=0
      KM(12)=0
      DO 1 I=1,7
      HH(I)=0.
    1 CONTINUE
      IFLAG=0
      NOC=0
      LOH=0
C**   FIRST OF ALL, SET INDICATOR TO ASSUME INVALIDITY
      DO 209 K=1,6
      JM(K)=0
  209 CONTINUE
C     JJ=NO. OF ITERATIONS
      DO 199 JJ=1,10
      TAA=1.-TA
      TBB=1.-TB
```

Appendix D

```
      TCC=1.-TC
      TDD=1.-TD
      TAS=TA**2
      TBS=TB**2
      TCS=TC**2
      TDS=TD**2
      TAAS=TAA**2
      TBBS=TBB**2
      TCCS=TCC**2
      TDDS=TDD**2
C
C     PROBABILITIES OF A RANDOMLY SELECTED SUBJECT BEING CLASSIFIED
C     INTO EACH OF 3X3 TABLE
C
      P20=P0*TBS*TDDS+P1*TAS*TDDS+P2*TBS*TCCS+PB*TAS*TCCS
      P21=2*P0*TBS*TD*TDD+2*P1*TAS*TD*TDD+2*P2*TBS*TC*TCC+
     12*PB*TAS*TC*TCC
      P22=P0*TBS*TDS+P1*TAS*TDS+P2*TBS*TCS+PB*TAS*TCS
      P10=2*(P0*TB*TBB*TDDS+P1*TA*TAA*TDDS+P2*TB*TBB*TCCS+
     1PB*TA*TAA*TCCS)
      P11=4.*(P0*TB*TBB*TD*TDD+P1*TA*TAA*TD*TDD+P2*TB*TBB*TC*TCC
     1+PB*TA*TAA*TC*TCC)
      P12=2.*(P0*TB*TBB*TDS+TA*P1*TAA*TDS+P2*TB*TBB*TCS+PB*TA*TAA*TCS)
      P00=P0*TBBS*TDDS+P1*TAAS*TDDS+P2*TBBS*TCCS+PB*TAAS*TCCS
      P01=2.*(P0*TBBS*TD*TDD+P1*TAAS*TD*TDD+P2*TBBS*TC*TCC+PB*TAAS*TC*TC
     1C)
      P02=P0*TBBS*TDS+P1*TAAS*TDS+P2*TBBS*TCS+PB*TAAS*TCS
      PP(1)=P20
      PP(2)=P21
      PP(3)=P22
      PP(4)=P10
      PP(5)=P11
      PP(6)=P12
      PP(7)=P00
      PP(8)=P01
      PP(9)=P02
      PZ2(JJ)=P02
C
      DO 143 I=1,9
      PP(I)=DABS(PP(I))
  143 CONTINUE
      DO 145 I=1,9
      IF(PP(I).LT.1.00D-10) IN(I)=1
  145 CONTINUE
      DO 10 I=1,9
      IF(IN(I).EQ.1) IN(10)=1
   10 CONTINUE
      IF (IN(10).EQ.1)GOTO 61
      GOTO 62
   61 WRITE(9,56)AST(J1),AST(8),K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1
     C,8),JJ
      WRITE(9,57) (PP(I),I=1,9)
      WRITE(9,58) (HH(I),I=1,7),D
      WRITE(9,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(9,32) (PZ2(LY),LY=1,10)
   32 FORMAT(' ','5',10X,5(D15.7),/,1X,11X,5(D15.7))
      IF(K1.LE.25) GOTO 16
      GOTO 11
   16 WRITE(6,56)AST(J1),AST(8),K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1
     C,8),JJ
   56 FORMAT('0','1',1X,A1,1X,A1,I4,9(1X,I3),1X,8(1X,F7.4),19X,I2)
      WRITE(6,57) (PP(I),I=1,9)
   57 FORMAT(' ','2',10X,9(D12.4,1X))
      WRITE(6,58) (HH(I),I=1,7),D
   58 FORMAT(' ','3',10X,7(D15.7),D12.4)
      WRITE(6,59)P0,P1,P2,PB,TA,TB,TC,TD
```

Appendix D

```
   59 FORMAT(' ','4',10X,8(D12.4,1X))
      WRITE(6,32) (PZ2(LY),LY=1,10)
   11 GOTO 239
C     PARTIAL DERIVATIVES
   62 P20P0=TBS*TDDS-TAS*TCCS
      P20P1=TAS*TDDS-TAS*TCCS
      P20P2=TBS*TCCS-TAS*TCCS
      P20TA=2.*(P1*TA*TDDS+PB*TA*TCCS)
      P20TB=2.*(P0*TB*TDDS+P2*TB*TCCS)
      P20TC=2.*(-P2*TBS+TC*P2*TBS-PB*TAS+TC*PB*TAS)
      P20TD=2.*(-P0*TBS+TD*P0*TBS-P1*TAS+TD*P1*TAS)
      P21P0=2.*(TBS*TD*TDD-TAS*TC*TCC)
      P21P1=2.*(TAS*TD*TDD-TAS*TC*TCC)
      P21P2=2.*(TBS*TC*TCC-TAS*TC*TCC)
      P21TA=4.*(P1*TA*TD*TDD+PB*TA*TC*TCC)
      P21TB=4.*(P0*TB*TD*TDD+P2*TB*TC*TCC)
      P21TC=2.*P2*TBS-4.*P2*TBS*TC+2.*PB*TAS-4.*PB*TAS*TC
      P21TD=2.*P0*TBS-4*P0*TBS*TD+2.*P1*TAS-4*P1*TD*TAS
      P22P0=TBS*TDS-TAS*TCS
      P22P1=TAS*TDS-TAS*TCS
      P22P2=TBS*TCS-TAS*TCS
      P22TA=2.*(P1*TDS+PB*TCS)*TA
      P22TB=2.*(P0*TDS+P2*TCS)*TB
      P22TC=2.*(P2*TBS+PB*TAS)*TC
      P22TD=2.*(P0*TBS+P1*TAS)*TD
      P10P0=2.*(TB*TBB*TDDS-TA*TAA*TCCS)
      P10P1=2.*TA*TAA*(TDDS-TCCS)
      P10P2=2.*(TB*TBB-TA*TAA)*TCCS
      P10TA=(2.*P1-4.*P1*TA)*TDDS+(2.*PB-4.*PB*TA)*TCCS
      P10TB=(2.*P0-4.*P0*TB)*TDDS+(2.*P2-4.*P2*TB)*TCCS
      P10TC=2.*(TBB*P2*TB+TAA*TA*PB)*(-2.+2.*TC)
      P10TD=2.*(P0*TB*TBB+P1*TA*TAA)*(-2.+2.*TD)
      P11P0=4.*(TB*TBB*TD*TDD-TA*TAA*TC*TCC)
      P11P1=4.*TA*TAA*(TD*TDD-TC*TCC)
      P11P2=4.*TC*TCC*(TB*TBB-TA*TAA)
      P11TA=(4.*P1-8.*P1*TA)*TD*TDD+(4.*PB-8.*PB*TA)*TC*TCC
      P11TB=(4.*P0-8.*P0*TB)*TD*TDD+(4.*P2-8.*P2*TB)*TC*TCC
      P11TC=4*(1.-2*TC)*(P2*TBB*TB+PB*TA*TAA)
      P11TD=4.*(1.-2.*TD)*(P0*TB*TBB+P1*TA*TAA)
      P00P1=TAAS*TDDS-TAAS*TCCS
      P00P0=TBBS*TDDS-TAAS*TCCS
      P00P2=TCCS*(TBBS-TAAS)
      P00TA=(-2.+2.*TA)*(P1*TDDS+PB*TCCS)
      P00TB=(-2.+2.*TB)*(P0*TDDS+P2*TCCS)
      P00TC=2.*(-1.+TC)*(P2*TBBS+PB*TAAS)
      P00TD=P0*(-2.+2.*TD)*TBBS+P1*(-2.+2.*TD)*TAAS
      P01P0=2.*(TD*TDD*TBBS-TC*TCC*TAAS)
      P01P1=2.*TAAS*(TD*TDD-TC*TCC)
      P01P2=2.*TC*TCC*(TBBS-TAAS)
      P01TA=(-4.*P1+4.*P1*TA)*TD*TDD+(-4.*PB+4.*PB*TA)*TC*TCC
      P01TB=(-4.*P0+4.*P0*TB)*TD*TDD+(-4.*P2+4.*P2*TB)*TC*TCC
      P01TC=2.*(1.-2.*TC)*(P2*TBBS+PB*TAAS)
      P01TD=2.*(1.-2.*TD)*(P0*TBBS+P1*TAAS)
      P12P0=2.*(TB*TBB*TDS-TA*TAA*TCS)
      P12P1=2.*TA*TAA*(TDS-TCS)
      P12P2=2.*(TB*TBB*TCS-TA*TAA*TCS)
      P12TA=(2.*P1-4.*P1*TA)*TDS+(2.*PB-4.*PB*TA)*TCS
      P12TB=(2.*P0-4.*P0*TB)*TDS+(2.*P2-4.*P2*TB)*TCS
      P12TC=4.*(P2*TB*TBB+PB*TA*TAA)*TC
      P12TD=4.*TD*(P0*TB*TBB+P1*TA*TAA)
      P02P0=TBBS*TDS-TAAS*TCS
      P02P1=TAAS*TDS-TAAS*TCS
      P02P2=TBBS*TCS-TAAS*TCS
      P02TA=(-2.*P1+2.*P1*TA)*TDS+(-2.*PB+2.*PB*TA)*TCS
      P02TB=(-2.*P0+2.*P0*TB)*TDS+(-2.*P2+2.*P2*TB)*TCS
      P02TC=2.*TC*(P2*TBBS+PB*TAAS)
```

Appendix D

```
      P02TD=2.*TD*(P0*TBBS+P1*TAAS)
C     COMPUTING EFFICIENCY SCORES
      S(1)=AF*P20P0/P20+AE*P21P0/P21+AD*P22P0/P22+BF*P10P0/P10+BE*P11P0
     1/P11+BD*P12P0/P12+CF*P00P0/P00+CE*P01P0/P01+CD*P02P0/P02
      S(2)=AF*P20P1/P20+AE*P21P1/P21+AD*P22P1/P22+BF*P10P1/P10+BE*P11P1/
     1P11+BD*P12P1/P12+CF*P00P1/P00+CE*P01P1/P01+CD*P02P1/P02
      S(3)=AF*P20P2/P20+AE*P21P2/P21+AD*P22P2/P22+BF*P10P2/P10+BE*P11P2/
     1P11+BD*P12P2/P12+CF*P00P2/P00+CE*P01P2/P01+CD*P02P2/P02
      S(4)=AF*P20TA/P20+AE*P21TA/P21+AD*P22TA/P22+BF*P10TA/P10+BE*P11TA/
     1P11+BD*P12TA/P12+CF*P00TA/P00+CE*P01TA/P01+CD*P02TA/P02
      S(5)=AF*P20TB/P20+AE*P21TB/P21+AD*P22TB/P22+BF*P10TB/P10+BE*P11TB/
     1P11+BD*P12TB/P12+CF*P00TB/P00+CE*P01TB/P01+CD*P02TB/P02
      S(6)=AF*P20TC/P20+AE*P21TC/P21+AD*P22TC/P22+BF*P10TC/P10+BE*P11TC
     1/P11+BD*P12TC/P12+CF*P00TC/P00+CE*P01TC/P01+CD*P02TC/P02
      S(7)=AF*P20TD/P20+AE*P21TD/P21+AD*P22TD/P22+BF*P10TD/P10+BE*P11TD
     1/P11+BD*P12TD/P12+CF*P00TD/P00+CE*P01TD/P01+CD*P02TD/P02
C     INFORMATION MATRIX ELEMENTS
      H(1,1)=P20P0**2/P20+P21P0**2/P21+P22P0**2/P22+P10P0**2/P10+P11P0**
     12/P11+P12P0**2/P12+P00P0**2/P00+P01P0**2/P01+P02P0**2/P02
      H(1,2)=P20P0*P20P1/P20+P21P0*P21P1/P21+P22P0*P22P1/P22+P10P0*P10P1
     1/P10+P12P0*P12P1/P12+P00P0*P00P1/P00+P01P0*P01P1/P01+P02P0*P02P1/
     2P02+P11P0*P11P1/P11
      H(1,3)=P20P0*P20P2/P20+P21P0*P21P2/P21+P22P0*P22P2/P22+P10P0*P10P2
     1/P10+P11P0*P11P2/P11+P12P0*P12P2/P12+P00P0*P00P2/P00+P01P0*P01P2/P
     201+P02P0*P02P2/P02
      H(1,4)=P20P0*P20TA/P20+P21P0*P21TA/P21+P22P0*P22TA/P22+P10P0*P10TA
     1/P10+P11P0*P11TA/P11+P12P0*P12TA/P12+P00P0*P00TA/P00+P01P0*P01TA/
     2P01+P02P0*P02TA/P02
      H(1,5)=P20P0*P20TB/P20+P21P0*P21TB/P21+P10P0*P10TB/P10+P11P0*P11TB
     1/P11+P12P0*P12TB/P12+P00P0*P00TB/P00+P01P0*P01TB/P01+P02P0*P02TB/
     2P02+P22P0*P22TB/P22
      H(1,6)=P20P0*P20TC/P20+P21P0*P21TC/P21+P10P0*P10TC/P10+P11P0*P11TC
     1/P11+P12P0*P12TC/P12+P00P0*P00TC/P00+P01P0*P01TC/P01+P02P0*P02TC/
     2P02+P22P0*P22TC/P22
      H(1,7)=P20P0*P20TD/P20+P21P0*P21TD/P21+P22P0*P22TD/P22+P10P0*P10TD
     1/P10+P12P0*P12TD/P12+P00P0*P00TD/P00+P01P0*P01TD/P01+P02P0*P02TD/
     2P02+P11P0*P11TD/P11
      H(2,2)=P20P1**2/P20+P21P1**2/P21+P22P1**2/P22+P10P1**2/P10+P11P1**2/
     12/P11+P12P1**2/P12+P00P1**2/P00+P01P1**2/P01+P02P1**2/P02
      H(2,3)=P20P1*P20P2/P20+P21P1*P21P2/P21+P22P1*P22P2/P22+P10P1*P10P2
     1/P10+P11P1*P11P2/P11+P12P1*P12P2/P12+P00P1*P00P2/P00+P01P1*P01P2/P
     201+P02P1*P02P2/P02
      H(2,4)=P20P1*P20TA/P20+P21P1*P21TA/P21+P22P1*P22TA/P22+P10P1*P10TA
     1/P10+P11P1*P11TA/P11+P12P1*P12TA/P12+P00P1*P00TA/P00+P01P1*P01TA/
     2P01+P02P1*P02TA/P02
      H(2,5)=P20P1*P20TB/P20+P21P1*P21TB/P21+P22P1*P22TB/P22+P10P1*P10TB
     1/P10+P11P1*P11TB/P11+P12P1*P12TB/P12+P00P1*P00TB/P00+P01P1*P01TB/P
     201+P02P1*P02TB/P02
      H(2,6)=P20P1*P20TC/P20+P21P1*P21TC/P21+P22P1*P22TC/P22+P10P1*P10TC
     1/P10+P11P1*P11TC/P11+P12P1*P12TC/P12+P00P1*P00TC/P00+P01P1*P01TC/
     2P01+P02P1*P02TC/P02
      H(2,7)=P20P1*P20TD/P20+P21P1*P21TD/P21+P22P1*P22TD/P22+P10P1*P10TD
     1/P10+P11P1*P11TD/P11+P12P1*P12TD/P12+P00P1*P00TD/P00+P01P1*P01TD/
     2P01+P02P1*P02TD/P02
      H(3,3)=P20P2**2/P20+P21P2**2/P21+P22P2**2/P22+P10P2**2/P10+P11P2**
     12/P11+P12P2**2/P12+P00P2**2/P00+P01P2**2/P01+P02P2**2/P02
      H(3,4)=P20P2*P20TA/P20+P21P2*P21TA/P21+P22P2*P22TA/P22+P10P2*P10TA
     1/P10+P11P2*P11TA/P11+P12P2*P12TA/P12+P00P2*P00TA/P00+P01P2*P01TA/
     2P01+P02P2*P02TA/P02
      H(3,5)=P20P2*P20TB/P20+P21P2*P21TB/P21+P22P2*P22TB/P22+P10P2*P10TB
     1/P10+P11P2*P11TB/P11+P12P2*P12TB/P12+P00P2*P00TB/P00+P01P2*P01TB
     2/P01+P02P2*P02TB/P02
      H(3,6)=P20P2*P20TC/P20+P21P2*P21TC/P21+P22P2*P22TC/P22+P10P2*P10TC
     1/P10+P11P2*P11TC/P11+P12P2*P12TC/P12+P00P2*P00TC/P00+P01P2*P01TC
     2/P01+P02P2*P02TC/P02
      H(3,7)=P20P2*P20TD/P20+P21P2*P21TD/P21+P22P2*P22TD/P22+P10P2*P10TD
```

Appendix D

```
1/P10+P11P2*P11TD/P11+P12P2*P12TD/P12+P00P2*P00TD/P00+P01P2*P01TD
2/P01+P02P2*P02TD/P02
 H(4,4)=P20TA**2/P20+P21TA**2/P21+P22TA**2/P22+P10TA**2/P10+P11TA**
12/P11+P12TA**2/P12+P00TA**2/P00+P01TA**2/P01+P02TA**2/P02
 H(4,5)=P20TA*P20TB/P20+P21TA*P21TB/P21+P22TA*P22TB/P22+P10TA*P10TB
1/P10+P11TA*P11TB/P11+P12TA*P12TB/P12+P00TA*P00TB/P00+P01TA*P01TB
2/P01+P02TA*P02TB/P02
 H(4,6)=P20TA*P20TC/P20+P21TA*P21TC/P21+P22TA*P22TC/P22+P10TA*P10TC
1/P10+P11TA*P11TC/P11+P12TA*P12TC/P12+P00TA*P00TC/P00+P01TA*P01TC
2/P01+P02TA*P02TC/P02
 H(4,7)=P20TA*P20TD/P20+P21TA*P21TD/P21+P22TA*P22TD/P22+P10TA*P10TD
1/P10+P11TA*P11TD/P11+P12TA*P12TD/P12+P00TA*P00TD/P00+P01TA*P01TD
2/P01+P02TA*P02TD/P02
 H(5,5)=P20TB**2/P20+P21TB**2/P21+P22TB**2/P22+P10TB**2/P10+P11TB**
12/P11+P12TB**2/P12+P00TB**2/P00+P01TB**2/P01+P02TB**2/P02
 H(5,6)=P20TB*P20TC/P20+P21TB*P21TC/P21+P22TB*P22TC/P22+P10TB*P10TC
1/P10+P11TB*P11TC/P11+P12TB*P12TC/P12+P00TB*P00TC/P00+P01TB*P01TC/
2P01+P02TB*P02TC/P02
 H(5,7)=P20TB*P20TD/P20+P21TB*P21TD/P21+P22TB*P22TD/P22+P10TB*P10TD
1/P10+P11TB*P11TD/P11+P12TB*P12TD/P12+P00TB*P00TD/P00+P01TB*P01TD
2/P01+P02TB*P02TD/P02
 H(6,6)=P20TC**2/P20+P21TC**2/P21+P22TC**2/P22+P10TC**2/P10+P11TC**
12/P11+P12TC**2/P12+P00TC**2/P00+P01TC**2/P01+P02TC**2/P02
 H(6,7)=P20TC*P20TD/P20+P21TC*P21TD/P21+P22TC*P22TD/P22+P10TC*P10TD
1/P10+P11TC*P11TD/P11+P12TC*P12TD/P12+P00TC*P00TD/P00+P01TC*P01TD/
2P01+P02TC*P02TD/P02
 H(7,7)=P20TD**2/P20+P21TD**2/P21+P22TD**2/P22+P10TD**2/P10+P11TD**
12/P11+P12TD**2/P12+P00TD**2/P00+P01TD**2/P01+P02TD**2/P02
  DO 99 J=1,5
  K=J+1
  DO 99 I=K,7
99 H(I,J)=H(J,I)
  H(7,6)=H(6,7)
  DO 80 I=1,7
  DO 80 J=1,7
80 H(I,J)=H(I,J)*K
  CALL MINV(H,7,D,L,M)
  DO 52 I=1,7
  HH(I)=H(I,I)
52 CONTINUE
  DO 500 I=1,7
  C(I,1)=0.
500 CONTINUE
  DO 510 I=1,7
  DO 510 J=1,7
  C(I,1)=C(I,1)+H(I,J)*ST(J,1)
510 CONTINUE
  P(1,1)=P0
  P(2,1)=P1
  P(3,1)=P2
  P(4,1)=TA
  P(5,1)=TB
  P(6,1)=TC
  P(7,1)=TD
  DO 101 J=1,7
  P(J,1)=P(J,1)+C(J,1)
101 CONTINUE
  P0=P(1,1)
  P1=P(2,1)
  P2=P(3,1)
  PB=1-(P(1,1)+P(2,1)+P(3,1))
  TA=P(4,1)
  TB=P(5,1)
  TC=P(6,1)
  TD=P(7,1)
  NY=0
```

POOR COPY

Appendix D

```
      DO 39 J=1,7
      PR(J)=DABS(P(J,1))
      IF(PR(J).GT.1.0D5)NY=1
   39 CONTINUE
      IF(NY.EQ.1)GOTO 40
      GOTO 42
   40 WRITE(9,69)AST(J1),K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1,8),JJ
   69 FORMAT('0','1',1X,A1,1X,'*',I4,9(1X,I3),1X,8(1X,F7.4),19X,I2)
      WRITE(9,57)(PP(I),I=1,9)
      WRITE(9,58)(HH(I),I=1,7),D
      WRITE(9,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(9,32)(PZ2(LY),LY=1,10)
      IF(K1.LE.25)GOTO 48
      GOTO 47
   48 WRITE(6,69)AST(J1),K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1,8),JJ
      WRITE(6,57)(PP(I),I=1,9)
      WRITE(6,58)(HH(I),I=1,7),D
      WRITE(6,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(6,32)(PZ2(LY),LY=1,10)
   47 IN(10)=0
      KN(9)=0
      KN(10)=0
      K10=0
      KS=0
      NOC=0
      LOH=0
      NOK=0
      GOTO 239
   42 IF(IFLAG.EQ.1)GOTO 45
      IF((P0.LT.-1.0).OR.(P0.GT.2.0))KN(1)=1
      IF((P1.LT.-1.0).OR.(P1.GT.2.0))KN(2)=1
      IF((P2.LT.-1.0).OR.(P2.GT.2.0))KN(3)=1
      IF((PB.LT.-1.0).OR.(PB.GT.2.0))KN(4)=1
      IF((TA.LT.-1.0).OR.(TA.GT.2.0))KN(5)=1
      IF((TB.LT.-1.0).OR.(TB.GT.2.0))KN(6)=1
      IF((TC.LT.-1.0).OR.(TC.GT.2.0))KN(7)=1
      IF((TD.LT.-1.0).OR.(TD.GT.2.0))KN(8)=1
      DO 20 I=1,8
      IF(KN(I).EQ.1)KN(9)=1
   20 CONTINUE
      IF(KN(9).EQ.1)GOTO 49
      GOTO 50
   49 IFLAG=1
      IF(KN(9).EQ.1) IND=AST(3)
      IF(KN(10).EQ.1) IND=AST(5)
      IF(K10.EQ.1) IND=AST(4)
      IF(KS.EQ.1) IND=AST(6)
      WRITE(9,56)AST(J1),IND,K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1,8)
     C,JJ
      WRITE(9,57)(PP(I),I=1,9)
      WRITE(9,58)(HH(I),I=1,7),D
      WRITE(9,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(9,32)(PZ2(LY),LY=1,10)
      IF(K1.LE.25) GOTO 26
      GOTO 29
   26 WRITE(6,56)AST(J1),IND,K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1,8)
     C,JJ
      WRITE(6,57)(PP(I),I=1,9)
      WRITE(6,58)(HH(I),I=1,7),D
      WRITE(6,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(6,32)(PZ2(LY),LY=1,10)
   29 IF(JJ.GT.5)GOTO 239
      GOTO 45
   50 DO 198 K=1,7
      A(K)=DABS(C(K,1))
      IF((A(K).GT..001).AND.(JJ.EQ.10))GOTO 25
```

Appendix D

```
      IF (A(K).GT..001) GO TO 45
  198 CONTINUE
      GOTO 506
   25 K10=1
      GOTO 49
  506 IF((P0.LT.-0.05).OR.(P0.GT.1.05)) KM(1)=1
      IF((P1.LT.-0.05).OR.(P1.GT.1.05)) KM(2)=1
      IF((P2.LT.-0.05).OR.(P2.GT.1.05)) KM(3)=1
      IF((PB.LT.-0.05).OR.(PB.GT.1.05)) KM(4)=1
      IF((TA.LT.-0.05).OR.(TA.GT.1.05)) KM(5)=1
      IF((TB.LT.-0.05).OR.(TB.GT.1.05)) KM(6)=1
      IF((TC.LT.-0.05).OR.(TC.GT.1.05)) KM(7)=1
      IF((TD.LT.-0.05).OR.(TD.GT.1.05)) KM(8)=1
      IF(H(3,3).LT.0.) KM(9)=1
      IF(P2.LT.-0.05)KM(11)=1
      IF(P2.GT.1.05)KM(12)=1
      DO 46 I=1,9
      IF(KM(I).EQ.1)KM(10)=1
   46 CONTINUE
      IF(KM(10).EQ.1)GOTO 49
      KS=1
      SQ=DSQRT(H(3,3))
      DO 208 K=1,6
      Z(K)=ZN(K)*SQ
      IF(P2.LT.Z(K)) JM(K)=1
  208 CONTINUE
      GOTO 49
   45 IF(JJ.EQ.5)GOTO 12
      GOTO 199
   12 WRITE(9,56)AST(J1),AST(7),K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1
     C,8),JJ
      WRITE(9,57)(PP(I),I=1,9)
      WRITE(9,58)(HH(I),I=1,7),D
      WRITE(9,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(9,87)(C(K,1),K=1,7)
      IF(K1.LE.25) GOTO 60
      GOTO 63
   60 WRITE(6,56)AST(J1),AST(7),K1,AF,AE,AD,BF,BE,BD,CF,CE,CD,(PI(I),I=1
     C,8),JJ
      WRITE(6,57)(PP(I),I=1,9)
      WRITE(6,58)(HH(I),I=1,7),D
      WRITE(6,59)P0,P1,P2,PB,TA,TB,TC,TD
      WRITE(6,87)(C(K,1),K=1,7)
   87 FORMAT(' ','5',10X,3(D12.4,1X),13X,4(D12.4,1X))
   63 DO 13 K=1,7
      IF(C(K,1).GT..001)GOTO 14
   13 CONTINUE
      GOTO 15
   14 NOC=1
      GOTO 17
   15 IF(H(3,3).LT.0.)GOTO 28
      DO 27 K=1,7
      A(K)=DABS(C(K,1))
      IF(A(K).GT..001) NOK=1
   27 CONTINUE
      SQ=DSQRT(H(3,3))
      Z1=1.96*SQ
      Z2=1.65*SQ
      IF(P2-Z1)18,18,19
   18 JV1=1
      GOTO 21
   19 JV1=0
   21 IF(P2-Z2)22,22,23
   22 JV2=1
      GOTO 17
   23 JV2=0
```

Appendix D

```
      GOTO 17
   28 LOH=1
   17 CONTINUE
      IF(NOC.EQ.1)GOTO 30
      GOTO 41
   30 JK1=JV1
      JK2=JV2
      GOTO 44
   41 IF(LOH.EQ.1)GOTO 43
      GOTO 44
   43 JH1=JV1
      JH2=JV2
   44 IF(IFLAG.EQ.1)GOTO 239
  199 CONTINUE
  239 RETURN
      END
C                                                                       MINV  10
C     .....MINV  20
C                                                                       MINV  30
C        SUBROUTINE MINV                                                MINV  40
C                                                                       MINV  50
C        PURPOSE                                                        MINV  60
C           INVERT A MATRIX                                             MINV  70
C                                                                       MINV  80
C        USAGE                                                          MINV  90
C           CALL MINV(A,N,D,L,M)                                        MINV 100
C                                                                       MINV 110
C        DESCRIPTION OF PARAMETERS                                      MINV 120
C           A - INPUT MATRIX, DESTROYED IN COMPUTATION AND REPLACED BY  MINV 130
C               RESULTANT INVERSE.                                      MINV 140
C           N - ORDER OF MATRIX A                                       MINV 150
C           D - RESULTANT DETERMINANT                                   MINV 160
C           L - WORK VECTOR OF LENGTH N                                 MINV 170
C           M - WORK VECTOR OF LENGTH N                                 MINV 180
C                                                                       MINV 190
C        REMARKS                                                        MINV 200
C           MATRIX A MUST BE A GENERAL MATRIX                           MINV 210
C                                                                       MINV 220
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED                  MINV 230
C           NONE                                                        MINV 240
C                                                                       MINV 250
C        METHOD                                                         MINV 260
C           THE STANDARD GAUSS-JORDAN METHOD IS USED. THE DETERMINANT   MINV 270
C           IS ALSO CALCULATED. A DETERMINANT OF ZERO INDICATES THAT    MINV 280
C           THE MATRIX IS SINGULAR.                                     MINV 290
C                                                                       MINV 300
C     .....MINV 310
C                                                                       MINV 320
      SUBROUTINE MINV(A,N,D,L,M)                                        MINV 330
      DIMENSION A(1),L(1),M(1)                                          MINV 340
C                                                                       MINV 350
C        .....MINV 360
C                                                                       MINV 370
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE  MINV 380
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION      MINV 390
C        STATEMENT WHICH FOLLOWS.                                       MINV 400
C                                                                       MINV 410
C     DOUBLE PRECISION A,D,BIGA,HOLD                                    MINV 420
C                                                                       MINV 430
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS    MINV 440
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS      MINV 450
C        ROUTINE.                                                       MINV 460
C                                                                       MINV 470
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO      MINV 480
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS.  ABS IN STATEMENT  MINV 490
C        10 MUST BE CHANGED TO DABS.                                    MINV 500
```

Appendix D

```
C                                                                       MINV 510
C       ------------------------------------------------------------    MINV 520
C                                                                       MINV 530
C       SEARCH FOR LARGEST ELEMENT                                      MINV 540
C                                                                       MINV 550
      D=1.0                                                             MINV 560
      NK=-N                                                             MINV 570
      DO 80 K=1,N                                                       MINV 580
      NK=NK+N                                                           MINV 590
      L(K)=K                                                            MINV 600
      M(K)=K                                                            MINV 610
      KK=NK+K                                                           MINV 620
      BIGA=A(KK)                                                        MINV 630
      DO 20 J=K,N                                                       MINV 640
      IZ=N*(J-1)                                                        MINV 650
      DO 20 I=K,N                                                       MINV 660
      IJ=IZ+I                                                           MINV 670
   10 IF(DABS(BIGA)-DABS(A(IJ))) 15,20,20                               MINV 680
   15 BIGA=A(IJ)                                                        MINV 690
      L(K)=I                                                            MINV 700
      M(K)=J                                                            MINV 710
   20 CONTINUE                                                          MINV 720
C                                                                       MINV 730
C       INTERCHANGE ROWS                                                MINV 740
C                                                                       MINV 750
      J=L(K)                                                            MINV 760
      IF(J-K) 35,35,25                                                  MINV 770
   25 KI=K-N                                                            MINV 780
      DO 30 I=1,N                                                       MINV 790
      KI=KI+N                                                           MINV 800
      HOLD=-A(KI)                                                       MINV 810
      JI=KI-K+J                                                         MINV 820
      A(KI)=A(JI)                                                       MINV 830
   30 A(JI) =HOLD                                                       MINV 840
C                                                                       MINV 850
C       INTERCHANGE COLUMNS                                             MINV 860
C                                                                       MINV 870
   35 I=M(K)                                                            MINV 880
      IF(I-K) 45,45,38                                                  MINV 890
   38 JP=N*(I-1)                                                        MINV 900
      DO 40 J=1,N                                                       MINV 910
      JK=NK+J                                                           MINV 920
      JI=JP+J                                                           MINV 930
      HOLD=-A(JK)                                                       MINV 940
      A(JK)=A(JI)                                                       MINV 950
   40 A(JI) =HOLD                                                       MINV 960
C                                                                       MINV 970
C       DIVIDE COLUMN BY MINUS PIVOT (VALUE OF PIVOT ELEMENT IS         MINV 980
C       CONTAINED IN BIGA)                                              MINV 990
C                                                                       MINV1000
   45 IF(BIGA) 48,46,48                                                 MINV1010
   46 D=0.0                                                             MINV1020
      RETURN                                                            MINV1030
   48 DO 55 I=1,N                                                       MINV1040
      IF(I-K) 50,55,50
   50 IK=NK+I                                                           MINV1060
      A(IK)=A(IK)/(-BIGA)                                               MINV1070
   55 CONTINUE                                                          MINV1080
C                                                                       MINV1090
C       REDUCE MATRIX                                                   MINV1100
C                                                                       MINV1110
      DO 65 I=1,N                                                       MINV1120
      IK=NK+I                                                           MINV1130
      HOLD=A(IK)                                                        MINV1140
      IJ=I-N                                                            MINV1150
      DO 65 J=1,N                                                       MINV1160
```

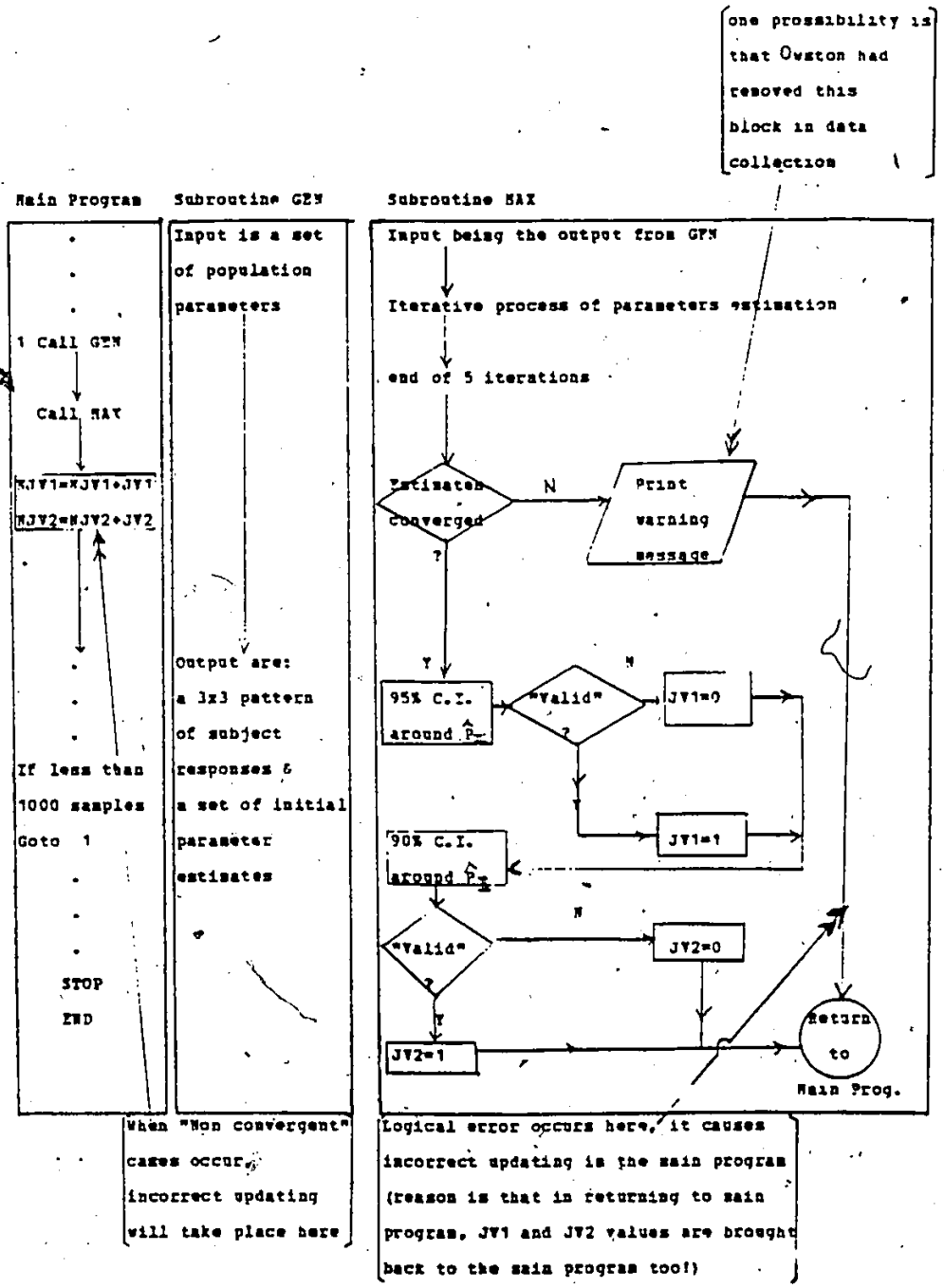
Appendix D

```
      IJ=IJ+N                                                    MINV1170
      IF(I-K) 60,65,60                                           MINV1180
   60 IF(J-K) 62,65,62                                           MINV1190
   62 KJ=IJ-I+K                                                  MINV1200
      A(IJ)=HOLD*A(KJ)+A(IJ)                                     MINV1210
   65 CONTINUE                                                   MINV1220
C                                                                MINV1230
C        DIVIDE POW BY PIVOT                                     MINV1240
C                                                                MINV1250
      KJ=K-N                                                     MINV1260
      DO 75 J=1,N                                                MINV1270
      KJ=KJ+N                                                    MINV1280
      IF(J-K) 70,75,70                                           MINV1290
   70 A(KJ)=A(KJ)/BIGA                                           MINV1300
   75 CONTINUE                                                   MINV1310
C                                                                MINV1320
C        PRODUCT OF PIVOTS                                       MINV1330
C                                                                MINV1340
      D=D*BIGA                                                   MINV1350
C                                                                MINV1360
C        REPLACE PIVOT BY RECIPROCAL                             MINV1370
C                                                                MINV1380
      A(KK)=1.0/BIGA                                             MINV1390
   80 CONTINUE                                                   MINV1400
C                                                                MINV1410
C        FINAL ROW AND COLUMN INTERCHANGE                        MINV1420
C                                                                MINV1430
      K=N                                                        MINV1440
  100 K=(K-1)                                                    MINV1450
      IF(K) 150,150,105                                          MINV1460
  105 I=L(K)                                                     MINV1470
      IF(I-K) 120,120,108                                        MINV1480
  108 JQ=N*(K-1)                                                 MINV1490
      JR=N*(I-1)                                                 MINV1500
      DO 110 J=1,N                                               MINV1510
      JK=JQ+J                                                    MINV1520
      HOLD=A(JK)                                                 MINV1530
      JI=JR+J                                                    MINV1540
      A(JK)=-A(JI)                                               MINV1550
  110 A(JI) =HOLD                                                MINV1560
  120 J=M(K)                                                     MINV1570
      IF(J-K) 100,100,125                                        MINV1580
  125 KI=K-N                                                     MINV1590
      DO 130 I=1,N                                               MINV1600
      KI=KI+K                                                    MINV1610
      HOLD=A(KI)                                                 MINV1620
      JI=KI-K+J                                                  MINV1630
      A(KI)=-A(JI)                                               MINV1640
  130 A(JI) =HOLD                                                MINV1650
      GO TO 100                                                  MINV1660
  150 RETURN                                                     MINV1670
      END                                                        MINV1680
      SUBROUTINE RANDU(IX,IY,YFL)
      IY=IX*65539
      IF(IY)5,6,6
    5 IY=IY+2147483647+1
    6 YFL=IY
      YFL=YFL*.4656613E-9
      RETURN
      END
```

POOR COPY

Appendix E

Flowchart to show a faulty logic in Owston's Fortran IV computer program



JV1, JV2 are variables to record decisions made by MAX-1 and MAX-2 on each sample.

NJV1, NJV2 are cumulative counters on "valid" indications for MAX-1 and MAX-2 respectively.

ABSTRACT

of

A Study of a Technique for

Validating Learning Hierarchies

A probabilistic technique for validating learning hierarchies was studied using a subset of an original set of hypothetical populations. This technique, called MAX, was developed by Owston. It was based upon the White & Clark Model of learning hierarchies and also was based on the maximum likelihood method of parameter estimation.

Various error types were identified in Owston's computer program in which he implemented the MAX technique. These error types were grouped into three categories. The first category is related to problems causing computer interruptions. The second category is associated with non-convergence of estimates. The third category is related to improperly converging estimates.

Two modified procedures, A and B, were developed. Procedure A was identical to the original procedure, except that the identified errors were considered. Procedure B was designed to improve upon procedure A. Ten rather than five

iterations were used, in procedure B, for the parameter estimation process; also, extra levels of confidence intervals were used.

The results obtained from using procedure A were compared with the corresponding results obtained by Owston. Differences in results were found, which cause one to question the conclusions made by Owston regarding the use of the MAX technique.

Results obtained from using procedure A were then compared with those obtained in procedure B. It was found that B was a better procedure for implementing the MAX technique.