DOCUMENT RESUME

ED 402 339 TM 025 924

AUTHOR	Scheuneman, Janice Dowd; And Others
TITLE	An Investigation of the Difficulty of Computer-Based Case Simulations.
PUB DATE	Apr 96
NOTE	29p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Case Studies; Computer Assisted Instruction; *Computer Simulation; Computer Software; *Difficulty Level; Medical Education; *Medical Students; *Prediction; Regression (Statistics); *Test Construction
IDENTIFIERS	*Case Method (Teaching Technique)

ABSTRACT

This study investigated the characteristics of Computer-based Case Simulations (CCS) that may be associated with case difficulty. Difficulty was defined as the average rating by physicians of examinee performance on a nine-point scale or the passing rate on the cases. Two data sets were used, one from an administration of 18 cases to 201 medical students, and the other from an administration of 22 cases to 117 students, with 13 cases being used on both occasions. Stepwise regression procedures were used separately for case properties and for analytic scoring key variables to identify the best predictors of case difficulty. Because of the small number of cases, regression results were evaluated for consistency across both data sets and both difficulty measures. For key variables, the best set of predictors included the number of different serious errors of commission, risk actions, and beneficial actions. In general, cases were more difficult for higher values of these variables. For case variables, the only consistent variable was the length of the paragraph that provided patient history, with longer paragraphs associated with more difficult cases. Other variables were less consistent, but were often related to the structure of the simulation or the severity of the patient condition. Although the findings for case variables were limited, the analyses were very helpful in illuminating the interconnections among the variables within cases. (Contains 7 tables and 15 references.) (Author/SLD)

An Investigation of the Difficulty of Computer-Based Case Simulations

Janice Dowd Scheuneman

Van Yihua Fan

Stephen G. Clyman

National Board of Medical Examiners

A paper presented at the Annual Meeting of the
National Council on Measurement in Education
April 1996

Abstract

This study investigated the characteristics of computer-based case simulations (CCS) that may be associated with case difficulty. Difficulty was defined as the average rating by physicians of examinee performance on a nine-point scale or the passing rate on the cases. Two data sets were used, one from an administration of 18 cases, the other from an administration of 22 cases with 13 cases used on both occasions. Stepwise regression procedures were used separately for case properties and for analytic scoring key varialbes to identify the best sets of predictors of case difficulty. Because of the small number of cases, regression results were evaluated for consistency across both data sets and both difficulty measures. For key variables, the best set of predictors included the number of different serious errors of commission, risk actions, and beneficial actions. In general, cases were more difficult for higher values of these variables. For case variables, the only consistent variable was the length of the paragraph that provided patient history, with longer paragraphs associated with more difficult cases. Other variables were less consistent, but were often related to the structure of the simulation or the severity of the patient condition. Although the findings for case variables were limited, the analyses were very helpful in illuminating the interconnections among the variables within cases.

Computer-based case simulations (CCS) are complex, unprompted, dynamic computer simulations of a patient-care environment. As a performance assessment instrument, they are intended to measure patient management skills in a realistic environment with simulated time and naturally unfolding clinical situations.

In CCS, the examinee is initially presented with a brief introduction to the patient's signs and symptoms. The examinee can request a history, conduct a physical examination, and write orders to perform diagnostic studies and initiate therapies. The patient's condition changes in response to the actions taken by the examinee and the time course of any underlying disease. The examinee needs to decide when, where, and how to care for the patient through this evolving course. Each action taken by the examinee is recorded by the computer, including canceled or refused actions, the simulated time of the action, and the cost.

As with most performance assessment instruments, each examinee must complete several cases for reliable measurement. At present, two types of CCS assessments are being field tested in medical colleges, an interdisciplinary instrument (iCCS) for use in the senior year of medical school and a discipline specific CCS (dCCS) in a field such as internal medicine or surgery for use in the junior year. A typical configuration includes eight to ten cases each averaging about 20 to 25 minutes to complete.

Planning is now underway to include CCS as one component of the United States Medical Licensing Examination. Before this plan is implemented, a number of technical and practical concerns will need to be resolved. The present study is designed to investigate which case properties are associated with case difficulty. Some of the potential advantages of better understanding the connection between case properties and difficulty of the case include the following:

1. Better instruction to case developers. Cases could be planned more efficiently; very easy or very difficult cases might be eliminated earlier in the developmental cycle.
2. Better targeting of cases to examinee ability. Experience has shown that cases that are too easy or too difficult for the examinee group are hard to score reliably.
3. Improved instructional feedback. Improved understanding of the nature of the problems provided to examinees would guide the nature of feedback provided to examinees and to medical school program directors.
4. Better control of the comparability of forms. Understanding case properties that affect difficulty would enable a more informed construction of alternate forms. Any efforts at statistical equating would also be facilitated to the extent that the case sets are parallel in both content and difficulty. Properties might actually be used in equating using methods like those suggested by Mislevy (1992).

Studies of Item Difficulty in Traditional Examinations

In psychometrics, item difficulty has been defined in terms of the performance of examinees, either in terms of the traditional p-value, percent correct responses by an examinee group, or in terms of a true score model such as the Rasch model, where difficulty measures are derived from examinee performance. Over the last few years, there has been a growing interest in going beyond this view to an understanding of what might be thought of as "intrinsic item difficulty," that is the difficulty that derives from the properties of the test item and the cognitive processes demanded from the examinee by the item.

The current interest in understanding the functioning of test items probably stems from earlier research in cognitive psychology that investigated components of laboratory tasks similar to those found on various tests of intelligence or aptitude, such as verbal or figural analogies (Carroll, 1976; Pellegrino & Glaser, 1979; and Sternberg, 1977a, 1977b). In these studies, theory-based task component were identified and mathematical models representing performance were developed. This work was generalized to actual test items that were similar to the laboratory tasks or to types of items where the components could be easily specified. The association of difficulty to components of tasks of such as geometric analogies, paper folding and hidden figures for which the link between the hypothesized underlying process and the item features was still fairly direct (Bejar & Yocum, 1986; Mulholland, Pellegrino, & Glaser, 1980; Smith & Green,1985; and Whitely & Schneider, 1981).

For many tests, however, the item content is complex and not easily represented objectively. Further, solution processes might differ in significant ways for different examinees. Models were derived empirically rather than from cognitive theory, using statistical methods to associate item properties with item difficulty. Component or complexity analyses of this type have been made with several types of items including reading comprehension, vocabulary, paragraph comprehension (reasoning) items, and GRE analytic items (Chalifour & Powers, 1988; Embretson & Wetzel, 1987; Scheuneman, Gerritz, and Embretson, 1989; Stenner, Smith, & Burdick, 1983).

Method

Data Source

The sample used in this study two groups of medical students tested from 1991 to 1994 as part of a series of special studies. The first group of 201 medical students received a total of 18 interdisciplinary cases administered in two separate sets. The second group of 117 students received 24 cases, in three separate sets of eight cases each in internal medicine, surgery, and pediatrics. These sets included 13 cases that were also administered to the first group. Two of the 24 cases were later dropped from analysis. Altogether, data were available for 27 different cases.

Development of Cases

Development of cases is a labor intensive effort involving many hours from staff and from physician volunteers. Case descriptions are developed by physicians according to pre-assigned specifications concerning the disease and certain patient characteristics. These descriptions are then reviewed by a case development committee that evaluates each case for its potential to be a good simulation. For example, does the case match its measurement objectives? Will sufficient opportunities be available for examinee intervention and for corrective feedback if inappropriate actions are taken? Is the task for the examinee well defined and can performance be interpreted unambiguously?

Cases that are approved by the case development committee are then prepared to be programmed for computer delivery. Flow charts are prepared in coordination with the physician author which specify the outcomes of different actions that might be taken by the examinee including screen notes that provide feedback on the developing condition of the patient. An untreated disease course, resulting from no action or inappropriate action, and times at which changes in the patient's conditions will occur within the different pathways are also specified. Once the case has been programmed, the case development committee will work through the case to see how it unfolds and further changes may be made at that time.

Finally, the case is presented to a key development committee. This committee of physicians determines the scoring points for possible actions that might be taken by the examinee. Actions are defined to be either beneficial, neutral, or not indicated. Actions that are not beneficial can be further divided into actions that are inappropriate but harmless; actions that present some risk to the patient; and flags, either of omission or commission, that represent serious errors in management that could severely jeopardize the patient.

Case Properties

A number of different variables were considered to represent case properties that might be associated with difficulty. Variables were identified through a review of the literature, the observation of physicians and staff involved in rating the problems, and inspection of the cases. Those finally selected could be categorized into several main areas. A list of the categories and variables in each is provided in Table 1.

Disease variables. The underlying diagnosis and the nature of the condition in this patient. Incidence frequencies for diagnoses and presenting conditions were obtained from the 1990 National Ambulatory Medical Care Survey.

Incidence rating. Two physicians rated on a three point scale the relative familiarity to medical students of cases like those presented in the problems. Ratings were averaged and results reviewed for internal consistency.

Incidence of diagnosis. Tables of incidence of the top 75 percent of most commonly diagnosed conditions were used to assign determine frequencies of most patient conditions. Because of the very large range of the frequencies and because some conditions were not listed, frequency categories were adopted to form five frequency categories with the lowest category made up by the unlisted conditions.

Incidence of presenting signs/symptoms. Tables of the top 75 percent most common complaints were used to obtain the incidence frequencies. Four frequency categories were formed with the lowest category again made up of unlisted conditions.

Area of Medicine. Cases in obstetrics/gynecology and in pediatrics were represented with dummy variables. Other cases were drawn from internal medicine, surgery, or emergency medicine, but these areas did not seem sufficiently distinct to code them separately.

Demographic variables. These variables describe the patient.

Age of patient

Gender of patient

Patient race. Race was not provided in the majority of the cases, although the patient was described as Black in a small number of cases in the dCCS data set.

Treatment setting variables. The setting in which the case management was initiated. Dummy variables were used to represent the *office*, *hospital ward*, or *emergency room* settings.

Patient condition variables.

Severity of Patient's Initial Condition. Three categories from mild to life threatening.

Acuity of Patient's Initial Condition. Three categories from chronic to acute.

Presence of Coexisting Conditions. In some cases patients have conditions other than those to be treated that may affect the management of the case.

Case management variables. These variables define the nature of the problem to be managed in this case and some of the examinee competencies the problem is designed to elicit.

Number of Problems to be Managed. Patients may have more than one problem or a less serious presenting problem may be masking a more serious one. Number of problems will be 1 or more.

Operative Intervention. Some cases require surgery as part of the successful management of the case.

Examinee Competencies. If the competency was specified to authors for case development, it was included here although other cases may have measured some aspects of these competencies as well. Competencies were selected from a much longer list if they did not overlap with other variables and were specified for at least five cases. Included as dummy variables were:

Recognizing subtle early signs/symptoms of a condition

Avoiding premature closure on possible diagnoses

Avoiding costly or invasive tests

Case structure variables. These variables concerned features of how the case was represented in the simulation.

Number of Treatment Pathways. Number of pathways specified in case flow chart.

Number of Screen Notes on Untreated Pathway. If the examinee fails to treat the underlying condition appropriately, screen notes will appear at intervals pointing out the deteriorating condition of the patient.

Patient History. At any point in the management of the case, the examinee may request the patient's history. A paragraph of background information is provided. The variables include the number of separate items of major and minor information included in the history, the number of words in the history, and the information density (number of information items divided by number of words). Major history items include those needed to arrive at a correct diagnosis. Minor history items are informative but not critical for forming a correct diagnosis. Also counted were the number of time intervals provided in developing the history of the present complaint.

Physical Examination Findings. If the examinee requests a physical examination, examination findings are provided. The number of these findings that suggest the correct diagnosis were counted as major physical examinations findings; informative but not critical findings were counted as minor physical examination findings.

Simulated Time. The length of the longest pathway, usually the untreated pathway, in simulated time.

Real Time Limit. The maximum amount of time the examinee is allowed to complete the case.

<u>Key variables</u>. These relate to the possible actions identified by the key development committee.

Benefits. Benefits could be subdivided in a number of ways. For this study the benefit variables were:

Number of more important and of less important benefits

Number of benefits that related to diagnosis, to treatment, and to monitoring

Percent of benefits that are diagnosis benefits

Number of benefits that need to be performed by a particular simulated time to receive maximum credit (time dependent benefits)

Percent of benefits that are time dependent

Number of time periods that have related time dependent benefits

Neutrals. Number of neutral actions specified in the scoring key. A large number of possible neutral actions suggests that ambiguity concerning the appropriate diagnosis may exist.

Inappropriates. Number of inappropriate actions specified in the scoring key. Again, many possible inappropriate actions suggest that examinees are likely to pursue incorrect diagnoses.

Risks. Number of risky actions specified in the scoring key.

Flags. Number of flags specified in the scoring key. Flags are both for important actions not taken (flags of omission) and for incorrect actions taken (flags of commission).

Values for neutrals, inappropriates, risks, and flags of commission were taken from field test data and reflect the actual performance of examinees.

Difficulty measures

Difficulty is not a well-defined concept in performance assessment measures such as CCS. Measures analogous to multiple choice difficulty measures might be developed from average examinee scores on each problem. At the time the research was initiated, however, analytic scoring procedures based on the key variables were still being investigated. We also wanted to consider the association of the key variables with difficulty, so that a difficulty measure derived independently of these variables would be desirable.

Fortunately, holistic scores were available. Both data sets were also used in evaluating different scoring procedures. As part of that effort, physician ratings of examinee performance were obtained for each case. The physicians were provided with transaction lists specifying the actions taken by the students in managing each case. Each transaction list was rated by two to six physicians on a holistic nine-point scale representing overall adequacy of the patient management. Definitions of the scale points were agreed upon by the raters. For example, a 1 meant the examinee missed a diagnosis of both primary and secondary conditions and did not treat the patient appropriately, while a 9 meant the examinee corretly diagnosed primary and secondary conditions, treated both, and provided followup (Clauser et al, 1995) . Later, a different committee was asked to rate each transaction list as a failing, borderline or passing performance. For study purposes, cases rated as borderline were considered passing (Clauser & Clyman, 1994).

The ratings were used to define difficulty for the cases. First was the average of the nine-point rating for each case, with higher scores reflecting easier cases. Second, was the percent of examinees passing each case with higher percents passing easier cases. Other definitions are possible, but two enabled us to determine if results were dependent on the difficulty measure chosen. The distribution of difficulty values, the mean, standard deviation, and range of the two difficulty measures for each data set is shown in Table 2. Correlations among the difficulty measures are shown in Table 3.

Results

Stepwise regression procedures were used to develop models for predicting difficulty from case properties. Because of the exploratory nature of the study, criteria for inclusion in the models was set high, p=.20 for inclusion in the model and .30 for removal, although for most analyses, the probability of the individual variables in the final model was less than .10. Care was also taken to avoid collinearity. If two highly correlated variables were included in the regression result, the analyses were repeated using only one of these variables on the list for stepwise selection. An exception was made if one of the variables was essentially uncorrelated with the difficulty measure and hence appeared to be functioning as a suppressor variable in the regression. The result of this procedure was that more than one apparently satisfactory solution was obtained for some of the difficulty measures.

Preliminary analyses suggested that the key variables be analyzed separately from the others. The key variables had higher intercorrelations among themselves and were generally more highly correlated with the difficulty measures than are the other case property variables. For some purposes, however, the key variables may be less useful since they are specified later in the case development process than are the other variables. Models were developed separately for the two data sets and for the two difficulty measures within each set. Since the number of cases is so small, the actual regression weights obtained by the model were of less interest that the variables found to be predictive of difficulty. Variables that appeared frequently in the different regression models for the four difficulty measures are likely to be those deserving more attention in future work.

Results for Key Variables

The correlations of the key variables with the difficulty measures is shown in Table 4. For the iCCS cases, slightly different models were obtained depending on whether the number of time-dependent benefits were included in the regression analysis. Different results were also obtained for the mean rating difficulty measure for the dCCS cases. The number of most important benefits (More Benefits in the table) was the best single predictor for this difficulty measure, but was also highly correlated with the other benefits variables, causing them to be excluded from the analyses. When the most important benefits variable was excluded, the second model was obtained. Only one model was obtained for the pass rate difficulty measure for the dCCS cases.

The results are shown in Table 5. Regression weights were considered unstable given the small number of cases; hence only the sign of the weight, and the probability (to give a rough indicator of importance) is provided in the table. A zero appears if the variable was used in the stepwise analysis, but did not appear in model. A blank indicates that the variable was not used in the analysis.

The most consistently important variable in these results is the risk actions, where more risk actions are associated with more difficult problems. Set against this are the flags in the models that these variables appear, with the presence of more flags of omission somewhat offsetting the difficulty effect of the risks and flags of commission adding to it. Benefits are clearly important in one form or another. The number or percent of timed benefits is associated with more case difficulty in six of the seven models. Two of the models show more diagnostic benefits associated with more difficult cases, with the effect somewhat offset by the number of treatment benefits. (Note that zero-order correlations with difficulty for flags of omission and treatment benefits are negative.)

Results of Case Property Variables

A total of 31 different variables were used in the regression analyses for each of the difficulty measures. Of these 29 were the same for iCCS and dCCS; however, only dCCS included more than one case with a Black patient and only iCCS included cases in obstetrics/gynecology. Only 16 of the 31 variables were included in any of the models developed from the regression analyses, although an additional seven variables affected the results for at least one of the difficulty measures and were removed from some of the analyses in order to obtain sensible models. The correlations of the difficulty measures with this subset of variables, those entering models and those that had to be withheld, are provided in Table 6.

The different regression analyses were complicated by the relationship of the various history variables to difficulty and the relatively high intercorrelations among the different history measures. However, reasonably sensible models were obtained only when the number of words in the history was included. For the difficulty rating measure for dCCS, three models were obtained depending on what other variables were included in the models. Only one result was obtained for the other difficulty measures. Results are shown in Table 7. Only the variables that appeared in one of these models are included in the tables; variables that were omitted from the analysis are provided in footnotes, or, if they appear in the table, are shown with an Om for omitted. NA indicates that the variable was not available for those cases.

The results show somewhat less consistency that do those for the key variables. Other than the appearance of the variable for the number of words in the history, the only variable to appear in the results for more than half the difficulty measures is the number of major physical examination findings, which has a different sign in different models. There are some consistencies of pattern, however. The most obvious is the larger number of variables associated with the case structure in all models. A less obvious pattern is the number of variables associated with the severity of the case.

The severity variable had low correlations with the difficulty measures and did not appear in any of the final models. Severity was highly correlated, however, with a number of variables that did appear in the models. Cases that were less severe tended to be initiated in the office, while more severe cases were initiated in the hospital ward or emergency room. Among the cases in this study, obstetrics/gynecology cases tended to be less severe and pediatric cases more severe. More severe cases were associated with more screen notes and more major findings in the history and physical examinations; less severe cases were associated with longer simulated times and longer real time limits, more minor findings in the history, and a concern with cost (measures costly/invasive competency). This suggests that the severity of the case is an important variable and should be considered in future work even though it did not appear in the final results.

Discussion and Conclusions

A study of the association of case difficulty with case properties has several potential benefits. Performance assessment problems, however, are relatively long so that very few problems are generally completedy by the same examinees. The large number of cases completed by the examinees in the data sets used in this study is rare. Nevertheless, the number of cases available was too small to produce definitive results. Efforts to overcome some of the effects of the small sample size included the use of more than two different difficulty measures and two independent data samples with overlapping problem sets.

Some consistency of results was found across the different analyses. For key variables, more possibilities for error as represented by a larger number of risks appears to be most important in the difficulty of the problems. Perhaps the more serious errors represented by flags of commission are easier for the examinee to avoid. A larger number of beneficial actions expected from the examinee also appears to lead to more difficult problems. Among other case properties, those variables associated with the problem structure and indirectly the severity of the patient's condition appear to most affect difficulty.

The association of more difficult problems with longer histories is intriguing as the number of words in the history was more predictive that the amount of information conveyed in the history, the importance of that information to the diagnosis, or the density of information (with low density implying the presence of extraneous or unneeded information in the history). This raises the possibility that the format in which the history is provided (currently as a paragraph of text) may affect how well the needed information is acquired by the examinee. The cases may be made easier by providing necessary points in an outline format, although this may reduce the fidelity of the case. In managing real cases, physicians must sort through the information provided by the patients to identify the important components.

Overall, although the statistical analyses were less informative than anticipated, the process of defining and coding the case variables and efforts to interpret the results provided a much greater understanding of the interconnections among the variables and of how the cases function. The knowledge gained from this study will be a rich source of hypotheses for future research.

References

Bejar, I. I., & Yocum, P. (1986). A generative approach to the development of hidden-figures items (RR-86-20-ONR). Princeton, NJ: Educational Testing Service.

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.

Chalifour, C., & Powers, D. E. (1988, May). Content Characteristics of GRE analytical reasoning items (RR 88-7). Princeton, NJ: Educational Testing Service.

Clauser, B. E., & Clyman, S. G. (1994). A contrasting groups approach to standard setting for performance assessments of clinical skills. Academic Medicine, 69 (October Supplement), S42-S44.

Clauser, B. E., Subhiyah, R. G., Nungester, R. J., Ripley, D. R., Clyman, S. G., & McKinley, D. (1995). Scoring a performance-based assessment by modeling the judgments of experts. Journal of Educational Measurement, 32, 397-415.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. Applied Psychological Measurement, 11, 175-193.

Mislevy, R. J., Sheehan, K. M.,& Wingersky, M. (1992). How to equate tests with little or no data (RR 92-20-ONR). Princeton, NJ: Educational Testing Service.

Mullholland, T., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. Cognitive Psychology, 12, 252-284.

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. In R. J. Sternberg & D. K. Detterman (Eds.), Human intelligence: Perspectives on its theory and measurement. Norwood, NJ: Ablex.

Scheuneman, J. D., Gerritz, K., & Embretson, S. E. (1991, July). Effects of prose complexity on achievement test item difficulty (RR 91-43). Princeton, NJ: Educational Testing Service.

Smith, R. M., & Green, K. E. (1985, April). Components of difficulty in paper-folding tests. Paper presented at the meeting of the American Educational Research Association, Chicago.

Stenner, A. J., Smith, M., & Burdick, D. S. (1983). Toward a theory of construct definition. Journal of Educational Measurement, 20, 305-316.

Sternberg, R. J. (1977a). Component processes in analogical reasoning. Psychological Review, 31, 356-378.

Sternberg, R. J. (1977b). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Erlbaum.

Whitely, S. E., & Schneider, L. M. (1981). Information structure for geometric analogies: A test theory approach. Applied Psychological Measurement, 5, 383-397.

Table 1

Description of Variables Representing Properties of Case Simulations

Disease Variables

Incidence of condition: Ratings
Incidence of condition: Frequency categories
Incidence of presenting signs/symptoms
Area of medicine: Obstetrics/gynecology; pediatrics; other

Demographic Variables

Age of patient
Gender of patient
Race of patient

Treatment Setting

Location of Initial encounter: Office, hospital ward, emergency room

Patient Condition Variables

Severity of patient's initial condition
Acuity of patient's initial condition
Presence of coexisting conditions

Case Management Variables

Number of patient conditions to be managed
Operative intervention required
Examinee competencies needed
- Recognizing subtle, early signs of condition
- Avoiding premature closure on diagnosis
- Avoiding costly and invasive therapies

Table 1 (continued)

Case Structure Variables

Number of treatment pathways
Number of screen notes on untreated pathway
Amount of information provided in patient history
Number of words in patient history
Longest simulated time allowed
Real time limit

Key Variables

Benefits

More important
Less important

Diagnosis
Treatment
Monitoring
Percent Diagnosis benefits

Time dependent
Percent time dependent
Number of time periods

Flags

Omission
Commission

Risks
Inappropriates
Neutrals

TABLE 2

FREQUENCY DISTRIBUTIONS AND SUMMARY STATISTICS FOR DIFFICULTY MEASURES

Rating			Percent Pass		
Rating	**iCCS**	**dCCS**	**Percent**	**iCCS**	**dCCS**
Above 6.5		2			
6.4-6.5		2	99-100		3
6.2-6.3		2	97-98	1	5
6.0-6.1		3	95-96	4	1
5.8-5.9		2	93-94	3	3
5.6-5.7		1	91-92	3	0
5.4-5.5		2	89-90	2	0
5.2-5.3		2	87-88	1	3
5.0-5.1	4	1	85-86	1	4
4.8-4.9	5	0	80-84	0	0
4.6-4.7	1	3	75-79	0	2
4.4-4.5	5	0	70-74	1	0
4.2-4.3	0	1	60-69	0	1
4.0-4.1	1	1	50-59	1	
Below 4.1	2		40-49	1	
Number of Cases	18	22		18	22
Mean	4.6	5.6		85.9	96.1
sd	0.6	0.8		15.6	9.1
Range	2.9-5.1	4.0-7.1		42-98	64-100

TABLE 3 CORRELATIONS AMONG DIFFICULTY MEASURES*				
	Rating i	**Pass i**	**Rating d**	**Pass d**
Rating iCCS	---	.87	.41	.91
Pass iCCS		---	.15	.79
%Rating dCCS			---	.66
%Pass dCCS				---

*Correlations between iCCS and dCCS measures are based on the 13 cases in common.

TABLE 4

CORRELATIONS OF KEY VARIABLES WITH DIFFICULTY MEASURES

	Rating iCCS	Rating dCCS	% Pass iCCS	% Pass dCCS
Benefits	-.28	-.55	-.14	-.41
More Imp.	-.36	-.57	-.23	-.47
Less Imp.	.29	.05	.24	.15
Diagnosis	-.13	-.44	-.10	-.40
Treatment	-.08	-.24	.23	-.09
Monitor	-.25	-.28	-.28	-.20
% Diag.	.08	-.03	-.04	-.14
Timed	-.46	-.29	-.38	-.43
% Timed	-.36	.38	-.33	-.11
Time Periods	.06	.19	.09	.13
Flags	-.01	-.40	.03	-.21
Omission	-.03	-.08	-.15	.05
Commission	.00	-.41	.10	-.27
Risks	-.61	-.30	-.53	-.61
Inapp.	-.08	.08	-.03	.14
Neutrals	-.24	.03	-.35	-.16

TABLE 5

SUMMARY OF REGRESSION MODELS USING KEY VARIABLES

	Rating iCCS		Rating dCCS		% Pass iCCS		% Pass dCCS
Variable	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	
Benefits	0				0		0
More Imp.	0			-**	0		0
Less Imp.	0		0	0	+		0
Diagnosis	0	0	-**		0	0	-**
Treatment	0	0	+		+**	+*	0
Monitor	0	0	+*		0	0	0
%Diag.	0		+**		0		0
Timed	-**				-**		0
%Timed	0	-**	+**	+**	0	-**	+
Time Periods	+	+*	0	0	0	0	0
Flags	0				0		0
Omission	+	0	+	0	+	0	+**
Com-mission	0	-*	0	0	0	-*	0
Risks	-**	-**	-**	-*	-*	-**	-**
Inapp.	0	0	0	0	0	+	0
Neutrals	-	-	0	+	0	-*	0
R^2	.706	.742	.674	.834	.741	.734	.754
#Variables	5	5	4	7	5	6	4
F	5.8	6.9	8.8	10.0	6.9	5.1	13.0
p	.006	.003	<.001	<.001	.003	.010	<.001

0= Does not enter model
* < .05
** <.01

TABLE 6

CORRELATIONS OF CASE PROPERTY VARIABLES WITH DIFFICULTY MEASURES

	Rating iCCS	Rating dCCS	% Pass iCCS	% Pass dCCS
Patient				
Age	.17	-.09	.01	.01
Gender	.01	.16	.16	.13
Race	NA	.03	NA	.01
Initial Location				
Emergency Room	.11	-.50	.29	-.06
Office	.10	.45	-.08	.20
Ward	-.22	.04	-.20	-.19
Incidence-Cat.	-.27	-.24	-.38	-.23
Pediatrics	-.31	-.03	-.16	-.01
OBGYN	.25	NA	.28	NA
Coexist. Prob.	.03	.06	-.11	-.15
Operative Int.	.07	.10	.07	.21
Premature Closure	-.48	-.07	-.41	-.14
Costly/Invasive	.02	.19	-.05	-.04
Longest Path	.12	.37	.06	.31
Time Limit	.05	-.27	.11	.02
Screennotes	-.27	-.52	-.03	- .35
History				
Words	-.59	-.42	-.44	-.64
Time	-.15	-.03	-.12	-.03
Total	.03	-.05	-.11	-.19
Density	.43	.34	.21	.38
Minor	-.05	.39	-.28	.00
Physical Exam				
Major	.13	-.27	.39	-.09
Minor	.12	.22	-.02	.24

TABLE 7
SUMMARY OF REGRESSION MODELS USING CASE VARIABLES

		Rating dCCS				
	Rating iCCS[1]	**Model 1[2]**	**Model 2[3]**	**Model 3[4]**	**% Pass iCCS[5]**	**% Pass dCCS[6]**
Patient Race	NA	-	-	-	NA	-*
Initial location						
Office					+**	Om
Emergency Rm			Om	-**		
Incidence - Cat.		-**	-*	Om		-**
Pediatrics						+
OBGYN	+	NA	NA	NA		NA
Coexist. Prob.		+*	+**	Om		
Operative Int			+		+**	
Premature Closure	-					
Longest Path			+*			+*
Time Limit		-*	-**	-		
Screennotes			-*			
History - words	-*	-**	-**	-	-**	-**
History - time						+*
Phys. Exam						
Major	+*	-**			+**	
Minor	+					
R^2	.731	.862	.910	.701	.877	.919
# Variables	5	6	8	4	4	6
F	3.8	10.4	10.1	7.0	14.3	19.0
p	.055	<.001	.002	.004	.001	<.001

* <.05
**<.01

Footnotes for Table 7

1 Regression did not include costly/invasive competency

2 Regression did not include age, gender, history--density, or costly/invasive competency

3 Regression did not include age, initial location--emergency room, or costly/invasive competency

4 Regression did not include gender, incidence--categories, or coexisting problems

5 Regression did not include history--minor or gender

6 Regression did not include history--total, history--density, or initial locations--hospital ward or office

NCME Annual Meeting, April 9-11, 1996

TM025924

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An Investigation of the Difficulty of Computer-Based Case Simulations	
Author(s): Janice Dowd Scheuneman, Van Yihua Fan, Stephen G. Clyman	
Corporate Source: National Board of Medical Examiners	Publication Date: April, 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[✓] **Sample sticker to be affixed to document**

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

_____ Sample _____

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

_____ Sample _____

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Janice Scheuneman	Position: Senior Evaluation Officer
Printed Name: Janice D. Scheuneman, PhD	Organization: National Board of Medical Examiners
Address: 3750 Market Street Philadelphia, Pa. 19104	Telephone Number: (215) 590-9669
	Date: June 3, 1996

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

March 12, 1996

Dear NCME Presenter,

Congratulations on being a presenter at NCME[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the NCME Conference. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (23)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

NCME 1996/ERIC Acquisitions
O'Boyle Hall, Room 210
The Catholic University of America
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the NCME web page (http://www.assessment.iupui.edu/ncme/ncme.html). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an NCME chair or discussant, please save this form for future use.