DOCUMENT RESUME

ED 476 927 TM 034 985

AUTHOR	Bergstrom, Betty A.; Blitz, David L.
TITLE	Job Task Analysis: An IRT Application.
PUB DATE	2000-04-15
NOTE	33p.; Paper presented at the Annual Meeting of the Society for Industrial and Organizational Psychology (15th, New Orleans, LA, April 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Item Response Theory; *Job Analysis; *Nurses; Nursing; *Rating Scales; Scaling

ABSTRACT

Job task analyses provide a link between job performance and examination content. This paper describes a methodology that uses Item Response Theory to place job task analysis data on an equal interval scale that allows for quantitative comparisons between tasks and provides a method for quantifying a test blueprint. An actual analysis was conducted using responses of 427 individuals in a nursing subspecialty who responded to a role delineation questionnaire to rate 161 nursing intervention tasks. Items with the highest calibrations on the latent variable received the highest relative percentage of test items, and items with the lowest calibrations received the lowest percentage of test items. Where job content must be closely tied to a predictive instrument for purposes of making informed personnel decisions, the rating scales derived from this method are more advantageous than those derived from raw scores. Others may adapt this methodology to create similar instruments to be used for purposes other than the one shown. (Contains 8 figures, 8 tables, and 22 references.) (Author/SLD)

Job Task Analysis: An IRT Application

Betty A. Bergstrom

David L. Blitz

Computer Adaptive Technologies, Inc.

Poster session presented April 15, 2000, at the Fifteenth Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, Louisiana. Correspondence concerning this paper should be directed to: David Blitz, Computer Adaptive Technologies, Inc., 1007 Church Street, 7th Floor, Evanston, IL 60201. Phone: (847) 866-2001; Fax: (847) 866-2002; e-mail: dblitz@catinc.com

Abstract

Job task analyses provide a link between job performance and examination content. This paper describes a methodology that utilizes Item Response Theory to place job task analysis data on an equal interval scale that allows for quantitative comparisons between tasks and provides a method for quantifying a test blueprint. An actual analysis was conducted; items with the highest calibrations on the latent variable received the highest relative percentage of test items, and items with the lowest calibrations received the lowest percentage of test items. Where job content must be closely tied to a predictive instrument for purposes of making informed personnel decisions, the rating scales derived from this method are more advantageous than ones derived from raw scores. Others may adapt this methodology to create similar instruments to be used for purposes other than the one shown.

Job Task Analysis: An IRT Application

Job task analyses, also known as audits of practice, practice analyses, task analysis inventories, or role delineation studies, are used to validate examinations by providing a link between job performance and examination content (AERA, APA, & NCME, 1985; EEOC, CSC, DOL, & DOJ, 1987; Guion, 1998; Kane, 1997; SIOP, 1987). Job task analyses help ensure that exam content specifications are current and relevant.

This paper proposes using an IRT rating scale analysis for job task analyses where the data on frequency and/or criticality of tasks or roles are collected using a Likert-type survey instrument. With this type of analysis, a linear scale is constructed that allows for consistent quantitative comparisons between tasks and provides a method for quantifying a test blueprint. The research presented herein extends previous work by one of the authors.

Theoretical Framework

The widespread use of job task analysis inventories today can be traced to research conducted by the U.S. Air Force (e.g., Morsh, 1964). Although originally used in the military, Harvey (1991, p.72) notes that "standardized job analysis questionnaires … have increased dramatically in number and popularity," particularly in the private sector. Perhaps one important reason for this method's increased popularity and use is that it lends itself nicely to job content validation (Gatewood & Feild, 1994).

Classical analysis of job task analysis data attempts to scale tasks on a hypothetical construct of interest (Crocker & Algina, 1986; Banerji, Smith & Dedrick, 1997). Tasks are placed on a continuum based on raw score averages obtained by arbitrarily assigning quantitative numbers to qualitative response categories (Not

Important=1, Somewhat important=2, Important=3, Extremely important=4). However, using raw score averages to quantify the latent construct is inadequate because of the non-linearity of the scale (Linacre, 1999; Michell, 1990). For numbers to represent amounts and enable quantitative comparisons, a linear scale must be developed in which the difference between respondents, the difference between items, and the difference between response categories can be compared.

The Rasch rating scale model (Wright & Masters, 1982) estimates the probability that a survey respondent will respond to a particular item with a particular response category as:

$$\log[P_{nij}/P_{ni(j-1)}] = B_n - D_i - F_j$$

where P_{nij} is the probability of respondent n scoring in category j of item i; $P_{ni(j-1)}$ is the probability of respondent n scoring in category j-1 of item i; B_n is the measure of respondent n; D_i is the difficulty of item i and F_j is the difficulty of step j. The unit of measure is the logit (log-odds unit) and the difficulty of an item is the logit value on the scale.

Rating scale categories are ordered "steps." If a respondent endorses an extreme statement, we expect her to endorse all less extreme statements. When four ordered levels are identified in an item, it can be thought of as having three steps (Figure 1):

Insert Figure 1 about here

A task on a role delineation study list can only be endorsed as "Important" if the first step from "Not Important" to "Somewhat Important" has also been endorsed. Thus the respondent who chooses "Important" can be considered to have chosen "Important" over "Somewhat Important" and "Not Important." For a three step item, the Rasch rating scale

models the probability of a respondent choosing a 4 rather than a 3, a 3 rather than a 2, a 2 rather than a 1.

While the rating scale model assumes a fixed set of rating points across items, it does not assume a fixed interval between the response categories. This is in contrast to using raw score points where the arithmetic difference between a 4 and a 3 is the same as the difference between a 3 and a 2. Using arbitrarily assigned raw scores assumes that the intervals between each of the qualitatively defined response categories (Not Important to Somewhat Important) are equal.

When the Rasch rating scale model is used to analyze job task data, a linear scale is constructed that enables actual quantitative comparisons between persons, items, and response categories. The difficulty of each item and the difficulty of each step are estimated on the same scale as the estimation of person measures.

Method

Data

Data consisted of responses from 427 individuals in a nursing subspecialty who responded to the role delineation study. One hundred sixty-one nursing intervention tasks were rated on frequency and criticality. The first scale allowed respondents to indicate how frequently they performed a specific nursing intervention and included the following response options:

A Several times per day
B About once a day
C About once per week
D About once per month
E Rarely, if at all

The criticality scale allowed respondents to indicate the importance of the intervention:

A Extremely important
B Important

C Somewhat important
D Not important

Procedure

The first step for the job task data analysis was to determine the sufficiency of the data to construct latent variables based on frequency of performance of an item, criticality of an item, or both. This was accomplished by analyzing the data using the Rasch rating scale model and the computer software program BIGSTEPS (Linacre, 1997). Two linear scales were constructed, representing the latent variables along which the items were placed based on their relative frequency or criticality.

After the linear scales were created, it was necessary to identify those traits that might be eliminated from consideration in constructing the desired latent variables. Tasks were ordered based on their frequency and criticality as determined by the item calibrations. Then the two sets of calibrations were plotted against each other. Items that were seen infrequently and/or were determined to be non-critical could be removed from consideration for the test blueprint. Items that exhibited misfit were also identified and reviewed for possible elimination from the test blueprint.

The data were examined for violations of the assumption of unidimensionality, a necessary step since Rasch rating scales are, by definition, unidimensional. Linacre (1998) has argued that following a Rasch analysis with a principal components analysis of the standardized residuals is preferred for determining multidimensionality. (The residual used here is the standardized difference between the observed response and the model expected response.) Thus, unrotated principal components analyses of the standardized residuals were conducted on the frequency and criticality responses, and resulting Eigenvalues were used to construct scree plots.

The final step was the construction of an examination blueprint identifying the content composition of the examination. A linear transformation of the frequency and/or criticality ratings converts the calibration of an item on the latent variable into a percentage of questions on an examination, so that items receiving a higher calibration are also assigned a higher percentage of questions on the test.

Results

Initial analysis of the frequency scale indicated that respondents were not using the entire five-point rating scale in a consistent manner (Figure 2). The inconsistency arose in the use of rating points 1 versus 2 and 3 versus 4. Basically, categories 3 and 4 were not being selected. The inconsistency was resolved by collapsing the five-point scale into a three-point scale in which 1 represented "daily," 2 represented "weekly or monthly," and 3 represented "rarely." Figure 3 shows the improved response probability curves after the scale was collapsed.

Insert Figure 2 about here

Insert Figure 3 about here

Analysis of the criticality scale indicated that respondents used the four-point rating scale in a satisfactory manner. However, they tended to use the "3" rating scale point, which represented "somewhat important" less frequently than other categories. They usually rated items as either "very important," "important," or "not important."

Figure 4 shows the frequency scale plotted against the criticality scale. Those tasks in the upper right quadrant are frequently seen and considered critical to practice.

Those tasks in the lower left quadrant are infrequently seen and not deemed critical to practice by the respondents.

Insert Figure 4 about here

The frequency variable spanned about a 6-logit range (-3 to +3) with negative calibrations representing interventions rarely practiced and positive calibrations representing interventions frequently practiced. The criticality latent variable spanned about 4 logits (-2 to +2) with negative calibrations representing unimportant items and positive calibrations representing extremely important items.

A scree plot (Figure 5) constructed from the principal components analysis of the frequency residuals indicated the presence of two additional factors, and another for the criticality residuals (Figure 6) indicated the presence of one additional factor. (While the Kaiser [1960] criterion would have led to the retention of additional factors for both principal components analyses, closer examination of the items within those additional factors failed to discern a meaningful pattern.) Closer inspection of the items comprising both first residual factors (frequency and criticality) indicated they were bipolar (refer to Stevens [1996] for a more detailed discussion), with a basic care orientation on the positive end and an education/counseling orientation on the negative end (Tables 1 – 4). The second residual factor (frequency only, accounting for 5% of the total variance) was also bipolar, and closer examination of items comprising this factor indicated a monitoring or measurement and interpretation orientation on the positive end and a self-care or nursing care orientation on the negative end (Tables 5-6). Although these factors

were present, they were determined to be facets within their respective dimensions rather than additional dimensions.

Insert Figure 5 about here

Insert Figure 6 about here

Insert Table 1 about here

Insert Table 2 about here

Insert Table 3 about here

Insert Table 4 about here

Insert Table 5 about here

Insert Table 6 about here

Subject matter experts reviewed these results. On the frequency and criticality scales, items having calibrations that fall below about -.50 on either or both scales are potential candidates for elimination from the examination blueprint. These are items that are infrequently encountered, unimportant, or both. In order to make this process more meaningful, subject matter expects reviewed criticality and frequency of tasks by content areas. They also reviewed items identified as misfitting. Figure 7 and Table 7 show the distribution of the content area "Perioperative care." Subject matter experts reviewed the data and decided to delete task 66, "Surgical Assistance," from consideration for the test

blueprint. This process was repeated for each content area. Following the removal of tasks, the remaining items were recalibrated.

Insert Figure 7 about here

Insert Table 7 about here

Once tasks that were considered to be "Not Important" and "Infrequently seen in Practice" were eliminated from consideration for the test blueprint and the remaining items were recalibrated, the decision had to be made regarding quantification of the test blueprint. The test blueprint could be derived from (a) the frequency scale only, (b) the criticality scale only, or (c) some combination of the frequency and criticality scales. In this particular case, subject matter experts decided to base the test blueprint on the frequency scale. The criticality scale was not used because of the high correlation between the frequency and the criticality scales.

Frequency calibrations are transformed to relative percentage of questions on the examination using a procedure developed by Lunz, Stahl and James (1989). This methodology allows for the development of two equations structured as linear transformation equations with two unknowns. The equations are solved for the slope and intercept constants:

$$Y = A + BX \tag{1}$$

where

Y = transformed percent of items from the Rasch calibrations of frequency
X = Rasch calibration for the frequency of each Nursing Intervention
A = Intercept constant
B = Slope constant

Once these two constants are known, a standard linear transformation formula is used to transform the item calibrations into relative percentages. Initially, the highest calibration is arbitrarily assigned 5% of the test and the lowest calibration is assigned 1%. The calibrations from this particular data appear in Table 8, below:

Insert Table 8 about here

The two simultaneous equations are set up as follows:

$$Y1 = A + X1B \quad \text{or} \quad .05 = A + 1.95B \tag{3}$$
$$Y2 = A + X2B \quad \text{or} \quad .01 = A - 1.65B \tag{4}$$

The two constants in the simultaneous equations are solved using the following equations:

$$A = (Y2 - Y1)/(X2 - X1) \tag{5}$$
$$B = Y1 - ((Y2 - Y1)/(X2 - X1)) * X1 \tag{6}$$

Applying these equations to the Nursing Intervention Classification Code Items yields:

$$A = .01111$$
$$B = .02833$$

This ensures that the item having the highest calibration on the latent variable receives the highest relative percentage of items on the test and the item having the lowest calibration receives the lowest percentage of items on the test.

Because of the large number of items on each variable, the total of the transformed percentages will generally not equal 100%. A correction factor is calculated and applied to each of the transformed percentages. The correction factor rescales the percentages in a way that maintains the relative percentages, but ensures that the total percentage on the test equals 100%.

$$R = 1 / \Sigma Y \tag{7}$$

R = Correction factor to be applied to the transformed percent of items
Y = Transformed percent of items from the Rasch calibrations of frequency

$\Sigma Y = 4.28$
$R = .23364$

Table 9 shows the final test blueprint for Nursing Interventions. Subject matter experts will use the test blueprint to balance their item bank and develop tests that reflect nursing practice.

Insert Table 9 about here

Discussion

Items are the operational definition of the latent trait. The more the items are spread across the range of person trait estimation, the better the definition of the trait variable. Using an IRT rating scale model places job task analysis data on a scale of equal interval units. Also, IRT rating scale models can account for missing or non-represented data (e.g., only "smart" respondents answer, etc). This is because IRT rating scale models generate item calibrations that are person- and test-free (refer to Wright & Masters [1982, p.6], or Lord [1980, pp.34-38] for a detailed discussion). And finally, when job task analysis data are scaled in equal interval units and missing data are taken into account, the test can be balanced to ensure that items with the highest calibrations on the latent variable receive the highest relative percentage of items on the test, and the items with the lowest calibrations receive the lowest percentage of items on the test. This property is not possible when raw scores alone are used to analyze the data.

These kinds of measurement improvements help strengthen confidence in the link between the predictor measure and the predictor construct in Figure 8 below (Binning

and Barrett [1989]; Guion [1998]). Binning and Barrett (1989) note that "...a single experiment cannot validate all four inferences simultaneously." However, they cite the contention that we can have confidence that a test validly measures a construct if "...(1) the domain of the other construct is well defined and (2) the assumption of a relationship between the two constructs is inarguable" (Nunnally & Berstein [1994, p.92], as originally appearing in Nunnally [1978]). The method described in this paper directly supports the first point. For the second point, we must rely upon subject matter experts. While not addressed in this paper, it is hoped that other efforts would be made to construct and validate a criterion measure, and then correlate it with the predictor measure, further strengthening the predictor-measure to criterion-construct link.

Insert Figure 8 about here

For the particular instrument described herein, plotting the criticality scale against the frequency scale allowed subject matter experts to determine what tasks are both critical and frequently seen, and enabled them to make informed decisions about content specifications for their certification examination. These decisions were carried into the test building process by translating the IRT calibrations into relative percentages of items on the exam. Thus, performing a job task analysis using the IRT rating scale model produced a test blueprint more representative of the profession studied. Others may choose to adapt this methodology to create similar instruments to be used for purposes other than licensure and certification (e.g., personnel selection). Where actual job content has to be tied as closely as possible to a predictive instrument for purposes of making informed personnel decisions, the method presented herein is more advantageous than previously used methods with rating scales derived using raw scores.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, D.C.: American Psychological Association.

Banerji, M., Smith, R. M., & Dedrick, R. F. (1997). Dimensionality of an early childhood scale: Using Rasch analysis and confirmatory factor analysis. Journal of Outcome Measurement, 1(1), 56-85.

Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74, 478-494.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: CBS College Publishing.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43(166), 38295-38309.

Gatewood, R.D., & Feild, H.S. (1994). Human resource selection (3rd ed.). Orlando, FL: Dryden Press.

Guion, R.M. (1998). Assessment, measurement, and prediction for personnel decisions. Mahwah, NJ: Lawrence Erlbaum Associates.

Harvey, R.J. (1991). Job analysis. In M.D. Dunnette & L.M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (pp.71-163). Palo Alto, CA: Consulting Psychologists Press.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kane, M. (1997). Model-based practice analysis and test specifications. Applied Measurement in Education, 10(1).

Linacre, J. M. (1997). BIGSTEPS [Computer Program]. Chicago, IL: MESA Press.

Linacre, J.M. (1998). Detecting multidimensionality: Which residual data-type works best? Journal of Outcome Measurement, 2(3), 266-283.

Linacre, J.M. (1999, April). Investigating rating scale category utility. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lunz, M.E., Stahl, J.A., & James, K. (1989). Content validity revisited: Transforming job analysis data into test specifications. Evaluation and the Health Professions. 12 (2). 192-206.

Michell, J. (1990). An introduction to the logic of psychological measurement. Hillsdale, NJ: Lawrence Erlbaum Associates.

Morsh, J.E. (1964). Job analysis in the United States Air Force. Personnel Psychology, 17, 717.

Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Nunnally, J.C. & Bernstein, I.H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

Society for Industrial and Organizational Psychology. (1987). Principles for the validation and use of personnel selection techniques (3rd ed.). College park, MD: Author.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago, IL: MESA Press.

Table 1. Frequency First Factor Positive Loadings

Factor Loading	Item Diff in logits	Entry Order	Item Content
0.62	0.95	70	11-Airway Suctioning
0.62	1.21	53	9-Cerebral Perfusion Promotion
0.60	0.77	71	11-Artificial Airway Managemen
0.58	1.12	52	9-Cerebral Edema Management
0.57	1.01	69	11-Airway Management
0.54	0.51	62	9-Tube Care: Ventriculostomy/L
0.54	0.63	75	11-Oxygen Therapy
0.53	0.44	74	11-Mechanical Ventilation
0.52	0.70	40	7-Electrolyte Management
0.51	0.73	76	11-Respiratory Monitoring
0.50	0.76	42	7-Fluid/Electrolyte Management
0.50	0.63	41	7-Electrolyte Monitoring
0.48	0.18	64	10-Postanesthesia Care
0.47	0.08	91	14-Shock Management: Vasogenic
0.47	0.23	38	7-Acid-Base Management
0.46	-0.06	90	14-Invasive Hemodynamic Monito
0.46	0.16	86	14-Dysrhthmia Management
0.45	0.52	58	9-Positioning: Neurologic
0.45	0.49	80	12-Pressure Ulcer Prevention
0.45	0.13	39	7-Acid-base Monitoring
0.44	0.57	55	9-Intracranial Pressure (ICP)
0.43	0.40	61	9-Subarachnoid Hemorrhage Prec
0.43	0.19	92	14-Shock Prevention
0.43	0.33	78	12-Pressure Management
0.42	0.32	88	14-Fluid Management
0.41	0.06	89	14-Fluid Resuscitation
0.41	0.26	68	11-Airway Insertion and Stabil
0.41	0.58	25	4-Enteral Tube Feeding
0.38	1.62	56	9-Neurologic Monitoring
0.38	1.15	72	11-Aspiration Precautions
0.37	-0.09	28	4-Total Parenteral Nutrition
0.36	0.68	59	9-Seizure Management
0.36	0.53	87	14-Embolus Precautions
0.36	0.97	20	3-Positioning
0.34	0.07	79	12-Pressure Ulcer Care
0.34	0.02	54	9-Dysreflexia Management
0.34	0.01	29	4-Tube Care: Gastrointestinal
0.33	-0.07	138	20-Physical Restraint
0.32	-0.12	84	13-Temperature Regulation
0.32	0.53	60	9-Seizure Precautions
0.31	-0.21	46	8-Conscious Sedation
0.31	0.69	19	3-Bed Rest Care
0.30	1.09	43	8-Analgesic Administration
0.30	0.16	85	14-Bleeding Precautions
0.30	0.84	130	19-Resuscitation

Table 2. Frequency First Factor Negative Loadings

Factor Loading	Item Diff in logits	Entry Order	Item Content
-0.58	-0.49	93	15-Activity Therapy
-0.57	-0.17	109	17-Coping Enhancement
-0.56	-0.16	110	17-Counseling
-0.53	-0.64	108	17-Body Image Enhancement
-0.53	-0.75	119	18-Teaching Group
-0.52	-0.64	94	15-Behavior Management
-0.51	-0.76	115	17-Support Group
-0.51	-0.19	160	24-Referral
-0.51	-0.17	116	17-Support System Enhancement
-0.50	0.10	117	18-Health Education
-0.49	-1.21	143	21-Home Maintenance Assistance
-0.48	-0.40	105	16-Communication Enhancement:
-0.46	-0.26	107	17-Anticipatory Guidance
-0.45	0.20	157	24-Health Care Information Exc
-0.45	0.11	140	21-Caregiver Support
-0.45	-0.59	106	16-Communication Enhancement:
-0.44	-0.94	145	21-Respite Care
-0.43	-0.13	122	18-Teaching: Prescribed Activi
-0.43	-1.04	95	15-Behavior Management: Overac
-0.42	0.33	120	18-Teaching: Individual
-0.41	-0.68	123	18-Teaching: Prescribed Diet
-0.41	0.02	139	20-Risk Identification
-0.41	0.28	124	18-Teaching: Prescribed Medica
-0.39	-1.47	112	17-Sexual Counseling
-0.39	-0.75	126	18-Teaching: Psychomotor Skill
-0.38	0.27	142	21-Family Involvement
-0.38	0.64	118	18-Teaching Disease Process
-0.38	-0.80	104	16-Communication Enhancement:
-0.38	0.09	159	24-Multidisciplinary Care Conf
-0.37	-0.21	128	19-Crisis Intervention
-0.37	-0.16	4	1-Exercise Therapy: Balance
-0.36	-1.28	127	18-Teaching: Sexuality
-0.36	-0.81	101	16-Memory Training
-0.35	-0.52	158	24-Health Policy Monitoring
-0.34	-0.25	153	23-Quality Monitoring
-0.32	-1.35	97	15-Behavior Management: Sexual
-0.32	-0.34	103	16-Communication Enhancement

Table 3. Criticality First Factor Positive Loadings

Factor Loading	Item Diff in logits	Entry Order	Item Content
0.75	0.66	75	11-Oxygen Therapy
0.69	0.31	70	11-Airway Suctioning
0.69	0.44	25	4-Enteral Tube Feeding
0.68	0.91	20	3-Positioning
0.65	0.56	76	11-Respiratory Monitoring
0.63	0.71	19	3-Bed Rest Care
0.63	0.47	12	2-Tube Care: Urinary
0.62	0.67	78	12-Pressure Management
0.62	0.12	71	11-Artificial Airway Managemen
0.59	0.32	69	11-Airway Management
0.58	0.63	88	14-Fluid Management
0.57	0.56	40	7-Electrolyte Management
0.56	0.59	72	11-Aspiration Precautions
0.54	0.60	80	12-Pressure Ulcer Prevention
0.53	0.59	42	7-Fluid/Electrolyte Management
0.51	0.21	38	7-Acid-Base Management
0.50	0.23	29	4-Tube Care: Gastrointestinal
0.50	-0.10	74	11-Mechanical Ventilation
0.50	0.17	138	20-Physical Restraint
0.48	0.21	39	7-Acid-base Monitoring
0.48	0.57	41	7-Electrolyte Monitoring
0.47	0.56	34	6-Oral Health Maintenance
0.46	0.49	24	4-Feeding
0.44	0.39	27	4-Nutrition Therapy
0.43	-0.61	92	14-Shock Prevention
0.43	0.59	155	23-Technology Management
0.42	-0.35	90	14-Invasive Hemodynamic Monito
0.42	0.05	86	14-Dysrhthmia Management
0.42	0.62	7	2-Bowel, Incontinence Care
0.41	0.40	32	6-Bathing
0.41	0.31	53	9-Cerebral Perfusion Promotion
0.40	-0.67	91	14-Shock Management: Vasogenic
0.40	0.32	84	13-Temperature Regulation
0.40	-0.36	28	4-Total Parenteral Nutrition
0.39	-0.27	89	14-Fluid Resuscitation
0.39	0.54	87	14-Embolus Precautions
0.39	-0.56	68	11-Airway Insertion and Stabil
0.38	0.30	81	12-Skin Care: Topical Treatmenent
0.38	-0.33	62	9-Tube Care: Ventriculostomy
0.37	0.21	52	9-Cerebral Edema Management
0.36	-0.17	61	9-Subarachnoid Hemorrhage Prec
0.36	0.04	82	13-Fever Treatment
0.34	-0.76	130	19-Resuscitation
0.34	1.17	43	8-Analgesic Administration
0.32	-0.25	26	4-Gastrointestinal Intubation
0.31	0.05	64	10-Postanesthesia Care
0.31	0.18	73	11-Cough Enhancement
0.30	-0.03	11	2-Diarrhea Management
0.30	0.56	77	12-Incision Site Care

Table 4. Criticality First Factor Negative Loadings

Factor Loading	Item Diff in logits	Entry Order	Item Content
-0.65	0.02	160	24-Referral
-0.65	0.41	124	18-Teaching: Prescribed Medica
-0.60	0.17	110	17-Counseling
-0.60	0.17	109	17-Coping Enhancement
-0.59	0.40	117	18-Health Education
-0.59	0.17	140	21-Caregiver Support
-0.58	0.33	157	24-Health Care Information Exc
-0.56	0.02	93	15-Activity Therapy
-0.55	-0.74	119	18-Teaching Group
-0.55	0.26	122	18-Teaching: Prescribed Activi
-0.53	-0.28	94	15-Behavior Management
-0.53	-0.82	115	17-Support Group
-0.52	0.52	120	18-Teaching: Individual
-0.52	-0.28	108	17-Body Image Enhancement
-0.50	-1.24	143	21-Home Maintenance Assistance
-0.49	0.88	118	18-Teaching Disease Process
-0.49	-0.05	107	17-Anticipatory Guidance
-0.48	-1.30	145	21-Respite Care
-0.48	0.13	116	17-Support System Enhancement
-0.46	-1.28	144	21-Normalization Promotion
-0.46	-0.02	159	24-Multidisciplinary Care Conf
-0.43	-0.39	128	19-Crisis Intervention
-0.42	-1.65	112	17-Sexual Counseling
-0.42	0.41	142	21-Family Involvement
-0.40	-0.64	95	15-Behavior Management: Overac
-0.40	0.49	125	18-Teaching: Procedure/Treatme
-0.39	-1.32	127	18-Teaching: Sexuality
-0.39	-0.29	158	24-Health Policy Monitoring
-0.38	-0.30	126	18-Teaching: Psychomotor Skill
-0.38	-1.07	141	21-Developmental Enhancement
-0.37	0.82	48	8-Medication Management
-0.37	-0.38	123	18-Teaching: Prescribed Diet
-0.34	-1.24	131	20-Abuse Protection: Elder
-0.34	-0.33	65	10-Preoperative Coordination
-0.33	-0.46	149	23-Peer Review
-0.31	-0.42	106	16-Communication Enhancement:
-0.31	0.30	4	1-Exercise Therapy: Balance
-0.30	-0.11	153	23-Quality Monitoring
-0.30	0.35	139	20-Risk Identification

Table 5. Frequency Second Factor Positive Loadings

Factor Loading	Item Diff in logits	Entry Order	Item Content
0.62	-0.35	90	14-Invasive Hemodynamic Monitor
0.55	-0.29	55	9-Intracranial Pressure (ICP)
0.55	-0.10	74	11-Mechanical Ventilation
0.48	-0.42	46	8-Conscious Sedation
0.47	-0.56	68	11-Airway Insertion and Stabil
0.44	-0.33	62	9-Tube Care: Ventriculostomy/L
0.42	-0.27	89	14-Fluid Resuscitation
0.41	-0.61	92	14-Shock Prevention
0.40	-0.67	91	14-Shock Management: Vasogenic
0.40	-0.76	130	19-Resuscitation
0.37	-0.17	61	9-Subarachnoid Hemorrhage Prec
0.36	0.31	53	9-Cerebral Perfusion Promotion
0.34	0.21	52	9-Cerebral Edema Management
0.34	-1.07	129	19-Organ Procurement
0.30	0.21	39	7-Acid-base Monitoring

Table 6. Frequency Second Factor Negative Loadings

Factor Loading	Item Diff in logits	Entry Order	Item Content
-0.57	0.29	17	2-Urinary Incontinence Care
-0.54	0.36	35	6-Self Care Assistence
-0.54	-0.44	16	2-Urinary Habit Training
-0.53	0.28	37	6-Swallowing Therapy
-0.53	0.00	15	2-Urinary Elimination Manageme
-0.53	-0.09	18	2-Urinary Retention Care
-0.51	-0.19	14	2-Urinary Catherization: Inter
-0.50	0.33	8	2-Bowel Management
-0.49	0.01	21	3-Positioning: Wheelchair
-0.47	-0.30	9	2-Bowel Training
-0.47	0.62	7	2-Bowel, Incontinence Care
-0.44	-0.46	13	2-Urinary Bladder Training
-0.40	0.62	136	20-Fall Prevention
-0.39	-0.10	33	6-Dressing
-0.38	0.30	4	1-Exercise Therapy: Balance
-0.37	0.49	24	4-Feeding
-0.37	0.60	80	12-Pressure Ulcer Prevention
-0.36	0.30	81	12-Skin Care: Topical Treatmen
-0.36	0.56	34	6-Oral Health Maintenance
-0.35	0.67	78	12-Pressure Management
-0.33	0.40	32	6-Bathing
-0.33	0.72	5	1-Exercise Therapy: Joint Mobi
-0.33	0.25	10	2-Constipation/Impaction Manag
-0.32	0.91	20	3-Positioning
-0.30	-0.33	134	20-Dementia Management
-0.30	0.06	103	16-Communication Enhancement

Table 7. Distribution of Content for Perioperative Care

Item Number	Nursing Interventions	Frequency Calibration	Criticality Calibration
64	Postanesthesia Care	0.05	0.22
65	Preoperative Coordination	-0.59	-0.35
66	Surgical Assistance	-1.88	-1.07
67	Surgical Preparation	-0.41	-0.35

Table 8. Values for Transforming Item Calibrations into Relative Percentages

Item	Rasch Calibration	Percent
Most Frequent	X1 = 1.95	Y1 =.05
Least Frequent	X2 = -1.65	Y2 =.01

Table 9. Nursing Interventions Test Blueprint[a]

Subarea	Percent
Elimination	8.01
Immobility	3.61
Nutrition	4.17
Physical Comfort	0.95
Self-care	4.36
Electrolyte & Acid/Base	3.83
Drug	4.22
Neurologic	8.10
Perioperative	1.82
Respiratory	6.46
Skin/Wound	3.80
Thermoregulation	1.72
Tissue Perfusion	5.08
Behavior	3.72
Cognitive	3.72
Coping	5.45
Patient	7.27
Crisis	1.35
Risk	5.56
Lifespan	2.80
Health Systems	0.72
Health Systems	5.66
Information	3.79
Total	100.15[a]

[a]Note: total exceeds 100% because of rounding.

Figure 1. Example of step ordering for a four-level rating scale.

1		2		3		4
Not Important		Somewhat Important		Important		Extremely Important
	Step 1		Step 2		Step 3	

Figure 2. Analysis of the initial five-point rating scale for frequency.

```
P      ÚÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀ¿
R  1.0 À111                                                   5555À
O      ³   11111111                                    5555555    ³
B      ³           111                             5555           ³
A      ³              11                         55               ³
B    .8 À               11                     55                 À
I      ³                  11                  5                   ³
L      ³                    1               5                     ³
I      ³                     1             5                      ³
T    .6 À                      1           5                       À
Y      ³                       1         5                        ³
    .5 À                       1        5                         À
O      ³                        1      5                          ³
F    .4 À                         1     5                          À
       ³                          1   5                           ³
R      ³                          1  5                            ³
E      ³                     2222221 5 444                        ³
S    .2 À                  222    33****3  4444                    À
P      ³               222     33  * *2 33     444                ³
O      ³           2222      33  4*   122 33      4444            ³
N      ³   22222222     3333344**5     11*2233333     44444444    ³
S    .0 À*********************55           11**********************À
E      ÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÀÙ
```

Figure 3. Analysis of the final three-point rating scale for frequency.

```
P      ÚÅÄÄÄÄÄÄÄÄÄÅÅÄÄÄÄÄÄÄÄÅÄÄÄÄÄÄÄÄÄÅÄÄÄÄÄÄÄÄÄÅÄÄÄÄÄÄÄÄÄÄÅÅÄÄÄÄÄÄÄÄÅ¿
R  1.0 Å1111111111                                          3333333333Å
O      ³          111111                              333333          ³
B      ³                111                        333                ³
A      ³                   11                    33                   ³
B   .8 Å                     1                  3                     Å
I      ³                      1                3                      ³
L      ³                       1              3                       ³
I      ³                        1            3                        ³
T   .6 Å                         1          3                         Å
Y      ³                          1        3                          ³
    .5 Å                           1      3                           Å
O      ³                           1     3                            ³
F   .4 Å                           2*222*2                            Å
       ³                          2  1 3  2                           ³
R      ³                        22    *    22                         ³
E      ³                       2     3 1     2                        ³
S   .2 Å                     22     3   1     22                      Å
P      ³                   22      3     1      22                    ³
O      ³                222      33       11      222                 ³
N      ³          222222      333           111       222222          ³
S   .0 Å**********333333333333                  111111111111**********Å
E      ÀÅÄÄÄÄÄÄÄÄÄÅÅÄÄÄÄÄÄÄÄÅÄÄÄÄÄÄÄÄÄÅÄÄÄÄÄÄÄÄÄÅÄÄÄÄÄÄÄÄÄÄÅÅÄÄÄÄÄÄÄÄÅÙ
       -6        -4        -2         0         2          4         6
```

Figure 4. Frequency vs. criticality calibrations for nursing interventions.

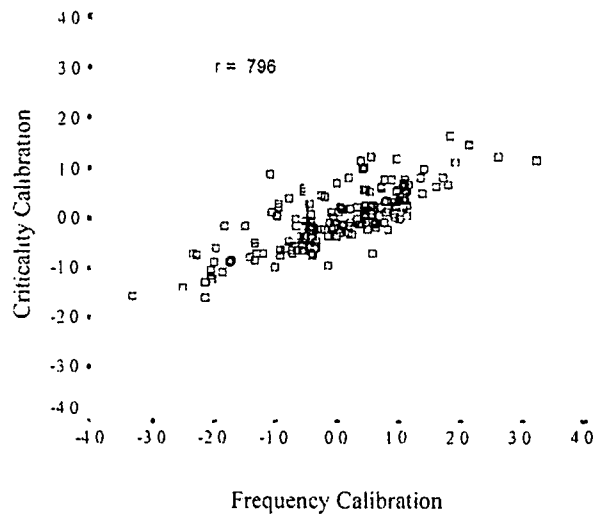


Figure 5. Scree plot of the principal components of standardized frequency residuals.

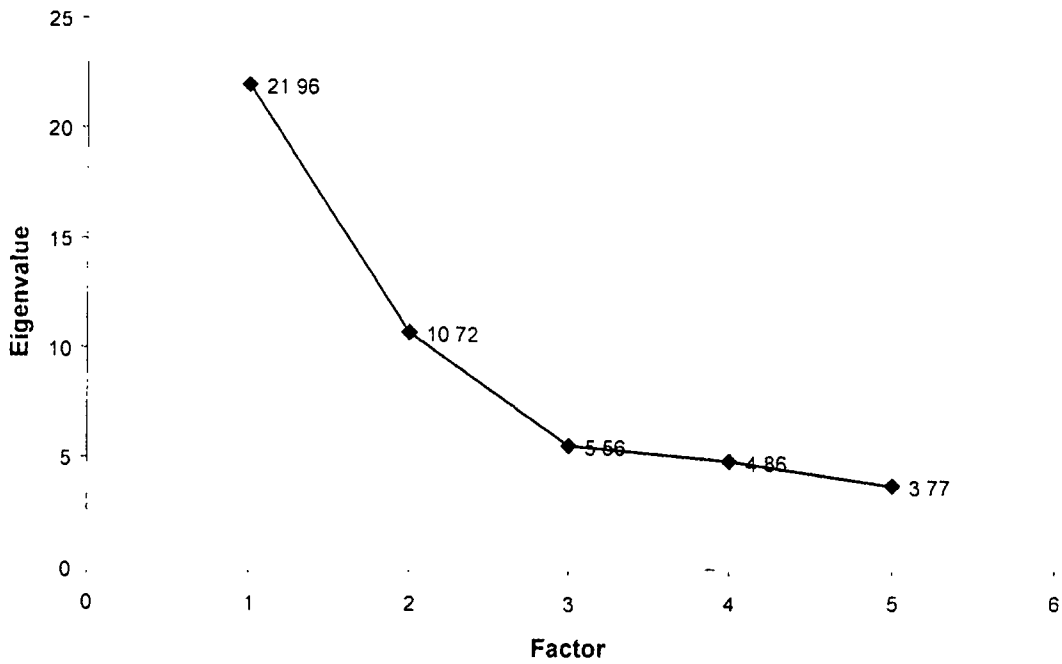


Figure 6. Scree plot of the principal components of standardized criticality residuals.

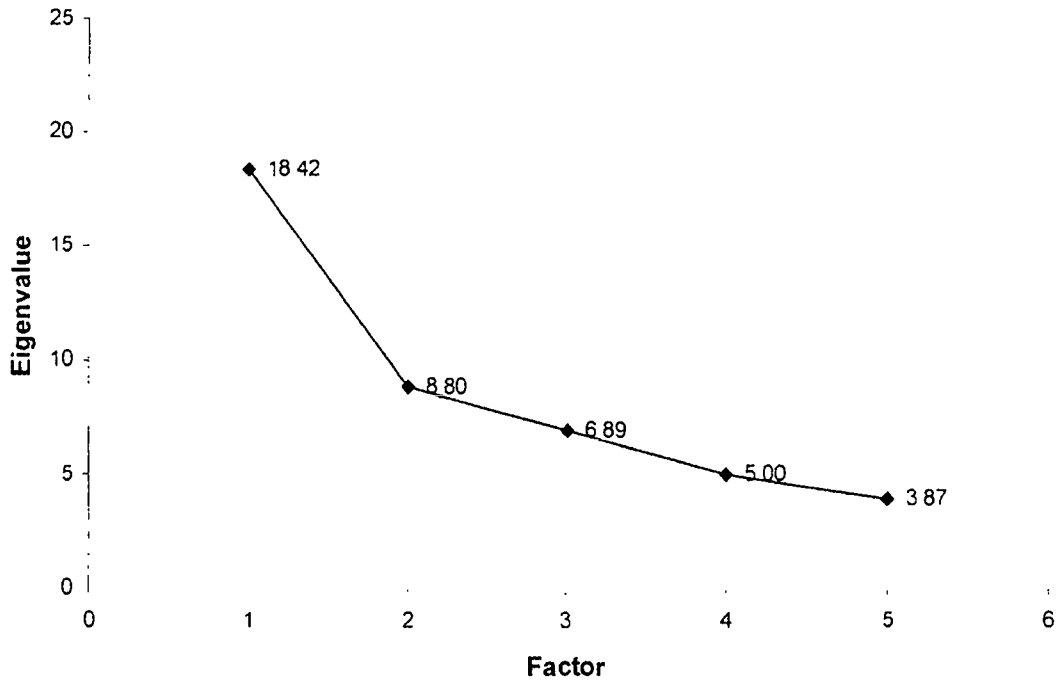


Figure 7. Frequency vs. criticality calibrations for perioperative care.

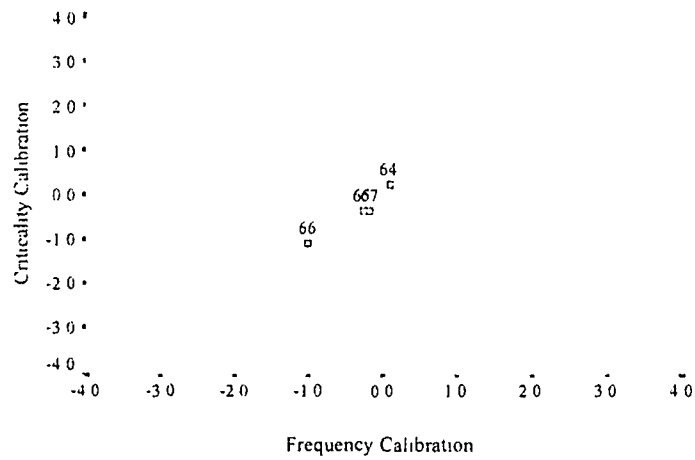
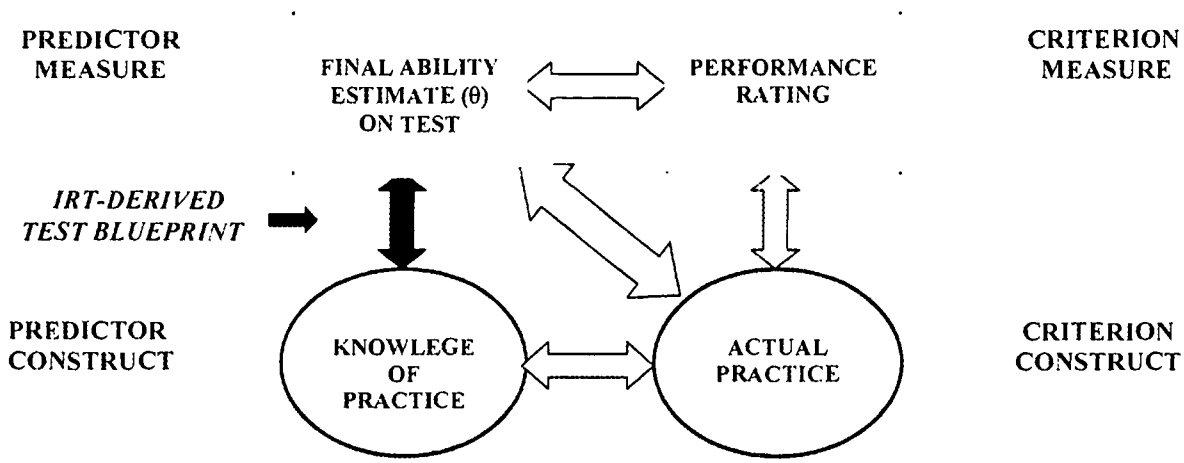


Figure 8. Basic linkages in the development of a predictive hypothesis. Adapted from Guion, R.M. (1998). Assessment, measurement, and prediction for personnel decisions. Mahwah, NJ: Lawrence Erlbaum Associates. Originally appearing in Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74, 478-494.



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM034985

I. DOCUMENT IDENTIFICATION:

Title: Job Task Analysis: An IRT Application	
Author(s): David L. Blitz and Betty A. Bergstrom	
Corporate Source:	Publication Date: April 2000

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 media, and sold through the ERIC Document Reproduction Service (EDRS). 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 and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

[X]

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

[ ]

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

[ ]

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic 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.

Sign here, please

Signature: David Blitz	Printed Name/Position/Title: David Blitz	
Organization/Address: Computer Adaptive Technologies, Inc. 1007 Church St. 7th Floor Evanston, IL 60201	Telephone: 847-866-2001	FAX: 847-866-2002
	E-Mail Address: dblitz@catinc.com	Date: 5/2003

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org