DOCUMENT RESUME

ED 476 428 TM 034 930

| | |
|---|---|
| AUTHOR | Dodeen, Hamzeh |
| TITLE | The Use of Person-Fit Statistics To Analyze Placement Tests. |
| PUB DATE | 2003-04-00 |
| NOTE | 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003). |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | EDRS Price MF01/PC01 Plus Postage. |
| DESCRIPTORS | Ability; *College Students; Foreign Countries; Goodness of Fit; Higher Education; *Statistical Analysis; *Student Placement; Test Use |
| IDENTIFIERS | *Person Fit Measures; United Arab Emirates |


ABSTRACT

Person fit is a statistical index that can be used as a direct measure to assess test accuracy by analyzing the response pattern of examinees and identifying those who misfit the testing model. This misfitting is a source of inaccuracy in estimating an individual's ability, and it decreases the expected criterion-related validity of the test being used. In placement tests, where individuals are usually classified based on their estimated ability levels, misfitting results in misclassification of individuals, which negatively affects the student as well as the academic organization. The study applied person-fit statistics as estimated by the standardized log-likelihood (l subscript z) index to analyze placement tests used by freshman college students (samples of 1,558 and 1,077). Results support the use of person-fit statistics to analyze placement tests effectively. The distribution of person-fit statistics on each test was very close to the standardized normal distribution, and both tests accurately assessed the intended student's ability. Results also show that ability level and person-fit statistics are not related. As for the aberrant response resulting from misfitting persons, two unusual common responses have been observed: missing most of the items at the end of the test and missing more easy items than expected at the beginning of the test. (Contains 2 tables and 23 references.) (SLD)



Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 476 428

# The Use of Person-Fit Statistics to Analyze Placement Tests

Hamzeh Dodeen, United Arab Emirates University

A paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.

Please send requests for copies to:

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
H. Dodeen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Hamzeh Dodeen
United Arab Emirates (UAE)
UAE University, Office of DVCAA
Al Ain, P. O. Box. 15551
(hdodeen@hotmail.com)

BEST COPY AVAILABLE

TM034930



ERIC

# The Use of Person-Fit Statistics to Analyze Placement Tests


## Abstract

Person fit is a statistical index that can be used as a direct measure to assess test accuracy by analyzing the response pattern of examinees and identifying those who misfit the testing model. This misfitting is a source of inaccuracy in estimating an individual's ability and it decreases the expected criterion-related validity of the test being used. In placement tests, where individuals are usually classified based on their estimated ability levels, misfiting results in misclassification of individuals which negatively affects the student as well as the academic organization. The study applied person-fit statistics as estimated by the standardized log-likelihood $(l_Z)$ index to analyze two placement tests used by freshmen college students. Results supported the use of person-fit statistics to effectively analyze placement tests. The distribution of person-fit statistics on each test was very close to the standardized normal distribution, and both tests accurately assessed the intended student's ability. Results also showed that ability level and person-fit statistics are not related. As for the aberrant responses resulting from misfitting persons, two unusual common responses have been observed: missing most of the items at the end of the test and missing more easy items than expected at the beginning of the test.

# The Use of Person-Fit Statistics to Analyze Placement Tests

## Introduction

Test scores are commonly used to assess an individual's ability. Whether the goal of the test is certification, college admission, detection of specific behavior, or personal selection, a decision about an individual's ability is usually made based on his or her test score. However, test scores might be an inaccurate measure of a person's ability; two individuals who score the same number correct on a test may have different aptitudes and abilities. The same score can be obtained by many probabilistic patterns (Harnisch & Linn, 1981). For example, if an individual answered the easiest 50% items of a test and a second individual answered the most difficult 50% items, the two would get the same ability estimation (total score). Both the individual and the organization or the academic institution is negatively affected by the inaccurate estimation of ability levels. An overestimate of an individual's ability could decrease the organizational productivity while an underestimate could exclude qualified individuals from possible opportunities; both could decrease the criterion-related validity of the test being used (Schmitt, Chan, Sacco, McFarland, & Jennings, 1999). One source of inaccuracy in estimating an individual's ability occurs when he/she misfits the test model (e.g., Rasch).

When item response theory (IRT) is used as a framework for testing, it is assumed that the test fits an IRT model. However, even when the test as a whole fits the testing model, some examinees do not. That is, taking into consideration their ability level, some examinees may give unusual or aberrant responses. One advantage of using IRT to represent test results is to be able to detect and identify misfitting persons.



ERIC Full Text Provided by ERIC

Statistically, a person misfits the model if there is a difference between his/her expected and observed response pattern. The expected response pattern for each person is determined based on both the ability level (θ) on the assessed trait and the assumed IRT model. For example, a low ability examinee who answered difficult items correctly would misfit most of the test models.

Person fit is a statistical index that can be used as a direct measure of assessment accuracy by analyzing response patterns and identifying persons with aberrant or unusual responses. It is useful to adequately understand and interpret the test results using test scores. Not only ability but also several psychological, cognitive, and personal factors affect an individual's responses to the test items (Hambleton, Swaminathan & Rogers, 1991). The effect of these factors may result in aberrant responses that are different than what is expected. In this context, person-fit statistics can be "a first step to trace persons whose answering behavior or part of it is the result of characteristics other than the latent ability that the test intends to measure" (Meijer & Sijtsma, 1994, p.7).

In general, aberrant or unusual responses are many; however, only the common ones have been identified or described in the literature. Meijer (1996) described some of them assuming a multiple-choice test with binary scoring. For example, sleeping behavior is observed when an examinee has trouble beginning the test (slow-to-warm-up); guessing behavior occurs when a low-ability person answers difficult items correctly by guessing blindly; cheating behavior can be the case when a low-ability student gets the most difficult items correct by copying the answers from a more competent neighbor; plodding behavior results from working slowly and not moving to the next item; and alignment error happens, for example, when a high-ability student skips an item in the



ERIC
Full Text Provided by ERIC

test but forgets to skip it in the answer sheet. Other aberrant responses might be due to fatigue, unfamiliarity with the topic or the test format (Swearingen, 1998), and scoring errors (Hulin, Drasgow, & Parsons, 1983).

If the test fits the model then it accurately represents the relationship between the ability level and the responses of the examinee. In other words, the test scores are accurate measures of an examinee's ability on the construct being assessed. Conversely, if the person misfits the model, then the accuracy of estimating his/her ability is negatively affected. Consequently, the decision made based on this assessment is inaccurate and this, as mentioned earlier, negatively affects the individual as well as the organization.

## Purpose

Placement tests are commonly used in classifying persons in different levels or categories based on their ability level on some construct(s). Taking into consideration the fact that placement tests are usually followed by a decision regarding individuals, and that an increasing number of persons are taking placement tests for different purposes, an effective and accurate procedure to evaluate the accuracy of assessment and classification in placement tests is necessary. Person fit statistics can be used to analyze response patterns and to identify persons with unusual responses. Hence, it can be used to assess accuracy of placement tests.

The purpose of this study was to investigate the use of person fit in evaluating the accuracy of placement tests. More specifically, the analysis and evaluation of a placement test was accomplished through addressing the following four research questions, with respect to the placement tests used:

1. What is the distribution of person-fit statistics?



ERIC

2. What is the percentage of misfitting persons to the assumed IRT-testing model?
3. Is there a relationship between person-fit statistics and ability level?
4. What are the common aberrant responses?

## Research on Person Fit

In the literature of person fit, few studies have been conducted in actual field applications. In this regard, Rudner (1995) stated that "although the need for person-fit statistics has been documented and uses for it have been suggested, for the most part, it has not yet been applied to many settings" (p.22). Meijer (1997) investigated the effect of model misfit on test validity in general and particularly on criterion-related validity. Meijer used simulated data that fit the three parameter logistic model (3PL) with four manipulated factors: number of items in the test, size of the correlation between the predictor test and the criterion test, proportion of nonfitting persons, and type of nonfitting. Results showed that the test validity decreased substantially when the type of misfit was severe or the correlation between the predictor and the criterion was high. A similar result was obtained when at least 15% of persons in the test misfit the model.

Yoes & Ho (1991) investigated the degree of person misfit on a national standardized achievement test. Three person-fit indicators (the unweighted standardized mean square, the standardized weighted mean square (INFIT), and the standardized likelihood index ($l_Z$)) were used to identify the percentage of misfitting persons. The study used three subtests (science, reading comprehension, and spelling) of the Stanford Achievement Test. Results indicated that the percentage of students misfitting the model (Rasch) was small. Results also supported the use of *lz* and INFIT indices in the detecting of aberrant responses as compared with the unweighted standardized mean square index.

Rudner (1995) used person-fit statistics in reporting and analyzing the results of the National Assessment of Educational Progress (NAEP). Using the IRT model-fit mean square statistics on data from the 1990 and 1992 NAEP assessment of mathematics, Rudner found that the data fit was good with very few abnormal responses (overall fit mean =.97, and standard deviation =.017). When blocking the analysis by states, fit was also good, and the same results were obtained for fit by race, community, and gender. In addition, no relationship was observed between person-ability and person-fit statistics (correlation ranges from -.20 to .17).

Fit statistics was applied in the investigation of a few specific educational issues, for example, gender and race differences and the use of different instruction methods. Frary (1982) applied four different fit statistics indices on a large sample of eighth grade students and found that males and whites are more likely to show aberrant responses than females and blacks respectively. Tatsuoka & Tatsuoka (1982) used two different methods of instruction to teach addition in signed-number operation to two equal groups of students and then calculated person-fit statistics for each. Results of this study showed that the difference between the two groups was significant. In another application, fit statistic was successfully used to identify schools that have curricula not matching the test content (Harnisch & Linn, 1981).

## Method

### Data

Person-fit analysis was applied on two placement tests used by freshmen college students in the United Arab Emirates University (UAEU). The tests are the English Placement Test (EPT) and the Arabic Placement Test (APT). Both tests are required for all freshmen students and conducted yearly to measure and classify students for levels of English or Arabic study. Data from the 2001-2002 administration of the EPT were used. The sample consisted of 1558 students (1214 females and 344 males) who responded to 119 multiple-choice items. The test had an internal reliability of .92. The data used for the APT was from the 2000-2001 administration. The test consisted of 90 multiple-choice items with an internal reliability of .84. The sample size was 1077 students (757 females and 320 males). Both tests were not speeded (more than 95% of students completed all items of each test).

<u>Person-Fit Index</u>

Person-fit indices are known by different names: "appropriateness measurement", "response aberrancy", " scalability", "individual consistency", and "norm conformity ". Some person-fit indices are person-group fit and others are person-model fit. Some indices are also based on the Classical Test Theory (CTT) while others are used under IRT models (see Meijer & Sijtsma, 2001 for extensive review of common person-fit indices). In the IRT framework, a person-fit index estimates the consistency of an individual's responses given his/her ability level (θ). Several person-fit indices have been derived and used. The standardized log-likelihood person-fit statistics $(l_Z)$ is one of the most accurate ones (Drasgow & Levine, 1986; Li & Olejnik, 1997; Nering, 1997; Nering & Meijer, 1998; Reise & Due, 1991; Yoes & Ho, 1991). $l_Z$ is the standardized estimate of $l_0$, which is given in the following formula (Schmitt, Cortina, & Whiteny 1993):

$$l_0 = \sum_{I=1}^{n} \{u_i \ln P_i(\theta) + (1-u_i)\ln[1-P_i(\theta)]\},$$

where $n$ is the number of items in the test, $u_i$ are the responses of the person to the $i$th item, and $P_i(\theta)$ is the probability of the response to item i given the person ability level θ. The $l_Z$ is given by standardizing $l_0$ as follows:

$$l_Z = \frac{l_0 - E(l_0)}{[Var(l_0)]^{1/2}},$$

where $E(l_0)$ is the expected value of $l_0$ and $Var(l_0)$ is the variance of $l_0$. The $E(l_0)$ is given in the following formula:

$$E(l_0) = \sum_{i=1}^{n}\{P_i(\theta)\ln P_i(\theta) + (1 - P_i(\theta))\ln[1 - P_i(\theta)]\},$$

while the $Var(l_0)$ is given by the formula:

$$Var(l_0) = \sum_{i=1}^{n} P_i(\theta)[1 - P_i(\theta)]\{\ln P_i(\theta)/[1 - P_i(\theta)]\}^2$$

According to Drasgow & Levine (1986), $l_Z$ has a distribution close to standardized normal distribution with a mean of 0 and a standard deviation of 1 at all θ levels. $l_Z$ is not recommended with short tests of fewer than 20 items (Reise & Due, 1991) unless high discriminating items are used (Meijer & Sijtsma, 1994). Large $l_Z$ values (greater than +2 or less than –2) are usually considered problematic responses (Ferrando & Lorenzo, 2000).

<u>Procedure</u>

A classical item analysis was conducted first on each test to determine the range of the discrimination index of items and the relative mean of items difficulty. This analysis is usually conducted to determine the appropriate IRT model to represent the data (Hambleton, et. al., 1991). Data from each test were calibrated using BILOG (Mislevy & Bock, 1990) which uses the marginal maximum likelihood algorithm. The number and percentage of misfitting items to the



ERIC Full Text Provided by ERIC

assumed model were used as another estimator of model-data fit. To clean up the data and to neutralize the effect of the item misfit on the analysis of person-fit, misfitting items were first identified and removed from the analysis. Items were then recalibrated and the item parameters (difficulty, discrimination, and guessing) and persons' ability parameter ($\theta$) were estimated and used in the WPerfit Program (Ferrando & Lorenzo, 2000) to calculate the person-fit $l_Z$ value for each person.

## Results

### Unidimensionality

This is the most important assumption underlying the use of IRT as a framework of testing. A test is said to be unidimensional if its items measure only one trait or ability. In practice, it is difficult to meet this assumption absolutely as many factors affect test performance. Factor analysis is commonly used to assess test unidimensionality. Reckase (1979) suggested that acceptable IRT parameter estimation could be obtained if the first factor accounts for at least 20% of the variance. In this study the assumption was checked through the factor analysis of the inter-item tetrachoiric correlations calculated by the program TESTFACT. It was found that the first factor for the EPT explained 23.5% of the variance and the second factor explained only 6.2%. For the APT, the first factor explained 28% while the second factor explained only 5.5% of the variance. Based on these results, it was reasonable to conclude that the unidimensionality assumption for the IRT models held for the two data sets.

### Model-data fit

Results of the classical item analysis for the two tests are shown in Table 1. The range of the item discrimination values for EPT and APT were .45 and .27 respectively. This large



ERIC

variation in the item discrimination values indicated that the 1PL model, which assumes equal discrimination values, was inappropriate to represent any of these data sets because it assumes equal discrimination values. In addition, as can be seen in Table 1, the averages of item difficulties for both tests were relatively low (.38 for EPT and .54 for APT). This means that the items were not relatively easy. If items are not easy, then guessing is likely to be assumed when answering questions especially multiple-choice ones (Hambleton, et. al., 1991). Since the 3PL considers both guessing and variation in discrimination values, it was more appropriate to represent these data sets.

---

Insert Table 1 about here

---

The last three columns of Table 1 show numbers and percentages of the misfitting items for each IRT model. These values which can be used as indicators of model-data fit. As was expected from the classical item analysis mentioned already, the 3PL was better than the other two models in representing the data sets. For example, using 1PL , 85 and 46 items misfit the model in EPT and APT respectively. The numbers of misfitting items were substantially decreased with the 3PL (23 and 12 items for the EAT and APT respectively). This result supported the use of the 3PL to represent the data of both tests.

Based on both the classical item analysis and the analysis of model-data fit, the 3PL was found to better fit the two data sets. To neutralize the effect of misfitting items on the analysis of person-fit statistics, all misfitting items in each test were deleted from the final calibration. This resulted in deleting 23 items from the EPT and 12 items from the APT.

In terms of the first research question, results showed that the distribution of person-fit statistics followed the familiar bell curve quite well with an expected mean of 0 and standard deviation of 1 (Table 2). The mean of person-fit statistics for the EPT was .17 with a standard deviation of .99, while the mean of that of the APT was very close to 0 (mean = .02) with a standard deviation of .92. The values of the Skewness and Kurtosis Tests were not significant ($p < .05$).

-----------------------------------------------------------------

Insert Table 2 about here

-----------------------------------------------------------------

The second research question addressed the percentage of the misfitting persons. The number and the percentage of missfitting persons on each test were very small. Only 75 (4.8%) and 33 (3.1%) persons misfit the test model in EPT and APT respectively. This result was less than what one might expect by chance only. The small number of misfitting persons indicated that the two placement tests had accurately assessed students' ability. If students were going to be classified based on these results, then the overwhelming majority of them would be accurately classified.

The third research question addressed the relationship between ability ($\theta$) and person-fit statistics. Because some of person-fit values were positive and some others were negative, person-fit absolute values were used in calculating the correlations. Results showed that person-fit statistics and ability were not correlated (correlation values were .002 and .13 for the EPT and APT respectively). In addition, the relationship between person-fit statistics and student college

GPA was checked and no correlation was found between the two variables (the correlations were -.01 and .02 for EPT and APT respectively).

The fourth research question addressed the identification of the common aberrant responses that appeared in each test. To save space, only the aberrant responses on the EPT were analyzed. Generally, giving an explanation for every specific unusual response is not a straightforward task. This difficulty is due to the fact that some different unusual behaviors can produce similar aberrant responses. For example, for a given test it may be difficult to decide whether some particular aberrant response is due to cheating or guessing behavior. Second, although the unusual behavior is likely to produce an unusual response, this is not always the case; for example, a person who guessed on a test and got some items correct but some others wrong might not misfit any IRT model (Meijer, & Sijtsma, 2001).

The responses of all misfitting persons (75 in the EPT) were analyzed first using the original order of the items as introduced in the test. Items were the rank-ordered based on their difficulty level, and the responses of the misfitting persons were analyzed again. In addition, all the aberrant responses were checked using the graphical plots offered by the WPerfit Program.

Two common unusual responses have been observed in the EPT. The first was missing most of the items as the end of the test (the last 30% items), more than what was expected based on the ability level. This unusual response was demonstrated by 19 students. The possible explanation to this kind of misfitting was fatigue. The test was long (119 items), which probably negatively affected students' motivation and/or concentration toward the end of the test. In addition, UAE students are not experienced enough with multiple-choice tests and their high school tests are usually much shorter. The second was missing more easy items than expected at



ERIC Full Text Provided by ERIC

the beginning of the test, but then answering correctly many more items than expected. Fifteen students demonstrated responses similar to this one. The possible explanation to this unusual response was either cheating or guessing. Given that these tests are administrated under strict conditions with a maximum penalty for cheating, guessing was very likely to be the cause of this response.

## Discussion

This study aims at applying the theoretically well-known person-fit statistics in analyzing placement tests. Person-fit statistics has been used as an accurate measure to identify aberrant responses that could be used as a measure of assessment accuracy. Placement tests are commonly used at different school levels and for different purposes. At the college level, placement tests are used to classify admitted students into studying levels. Misclassification of students based on the results of placement tests negatively affects both students and colleges. Taking into consideration the wide use of placement tests and the large number of examinees who take such tests, the accuracy of the assessment of these tests becomes a significant issue.

In both placements tests used in this study, the distribution of person-fit statistics was very close to the standardized normal distribution. This result supports the use of person-fit statistics to effectively analyze placement tests. This analysis includes evaluating the assessment accuracy of a test and identifying the aberrant responses. Evaluating the assessment accuracy can result in improving the test itself (e.g., using different formatting, changing the test length, or using another assessment tool). For example, if it is observed that blind guessing mainly causes person misfit, then using item formatting that reduces guessing (e.g., essay questions) can be a good alternative. Another example is shortening the test if it is found that fatigue causes the misfit. As for aberrant responses, the identification of the misfitting persons, their number, and percentage allows the organization to estimate the "size" and the effect of inaccurate estimation.

With regard to the relationship between ability and person-fit statistics, results show that there is no relationship between the two variables. This result is consistent with the result of Rudner (1995). The absence of a relationship between ability and person-fit indicates that there are psychological and personal factors other than ability responsible for misfitting. For example, level of motivation, level of test anxiety, tendency to guess (Hambleton, et. al., 1991), tendency to cheat, attitude toward the topic, level of attentiveness, and sincerity (Green, 1996) affect examinees' responses. This supports the use of person-fit statistics to identify examinees affected by such factors as it cannot be achieved using test scores. However, how such psychological and personal factors affect person-fit statistics has not been investigated yet. Further research needs to be directed toward studying the effect of these factors on person-fit statistics.

Identifying the number or the percentage of misfitting persons is essential in the evaluation of a test and its accuracy. However, a comprehensive understanding and evaluation of the test may not be accomplished without identifying the common aberrant responses demonstrated by misfitting persons and the best interpretation of each.



ERIC

References

Drasgow, F., & Levine, M. V. (1986). Optimal detection of certain forms of inappropriate test scores. Applied Psychological Measurement, 10, 59-67.

Ferrando, P. J., & Lorenzo, U. (2000). WPerfit: A program for computing parametric person-fit statistics and plotting person response curves. Educational and Psychological Measurement, 60(3), 479-487.

Frary, R. B. (1982). A comparison of person-fit measures. Paper presented to the annual meeting of the American Educational Research Association, New York, NY.

Green, K. E. (1996). The use of person fit statistics in mail surveys. Paper presented to the annual meeting of the American Educational Research Association, New York, NY.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage Publications.

Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 18,133-146.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory. Homewood, IL: Dow Jones-Irwin.

Li, F. M., & Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual responses patterns. Applied Psychological Measurement, 21(3), 251-231.

Meijer, R. R. (1996). Person-fit research: An introduction. Applied Measurement In Education, 9(1), 3-8.

Meijer, R. R. (1997). Person fit and criterion-related validity: An extension of the Schmitt, Cortina, and Whitney study. Applied Psychological Measurement, 21(2), 99-113.

Meijer, R. R., & Sijtsma, K. (1994), Detection of aberrant item score patterns: A review of recent developments. (ERIC Document Reproduction Service No. ED 389 751).

Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. Applied Psychological Measurement, 25(2), 107-135.

Mislevy, R. J., & Bock, R. D. (1990). BILOG maximum likelihood item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Software.

Nering, M. (1997). The distribution of person fit using true and estimated person parameter. Applied Psychological Measurement, 19(3),121-129.

Nering, M. L., & Meijer, R. R. (1998). A comparison of person response function and $l_Z$ person-fit statistics. Applied Psychological Measurement, 22(1), 53-69.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implication. Journal of Educational Statistics, 4, 207-230.

Reise, S. P., & Due, A. M. (1991). The influence of test characteristics on the detection of aberrant response patterns. Applied Psychological Measurement, 15, 217-226.

Runder, L. (1995). Analyzing national assessment of educational progress results. (ERIC Document Reproduction Service No. ED 379 346).

Schmitt, N., Chan, D., Sacco, J., McFarland, L., & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. Applied Psychological Measurement, 23(1), 41-53.

Schmitt, N., Cortina, J. M., & Whitney, D. J. (1993). Appropriateness fit and criterion-related validity. Applied Psychological Measurement, 17(2), 143-150.




ERIC

Swearingen, D. L. (1998). Person fit and its relationship with other measures of response set. Paper presented to the annual meeting of the American Educational Research Association, San Diego, CA.

Tatsuoka, K. K., & Tatsuoka, M. M. (1982). Detecting of aberrant responses patterns and their effect on dimensionality. Journal of Educational Statistics, 7(3), 215-231.

Yoes, M. E., & Ho, K. T. (1991). The degree of person misfit on a nationally standardized achievement test. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Table 1

Classical Item Analysis of EPT and APT and Misfitting Items

| Test | No. of items | Mean of item discrimination | Range of item discrimination | Mean of items difficulty | Number and percentage of misfitting items | | |
|---|---|---|---|---|---|---|---|
| | | | | | 1PL | 2PL | 3PL |
| EPT | 119 | .39 | .45 | .38 | 85 (71%) | 53 (45%) | 23 (19%) |
| APT | 90 | .26 | .27 | .54 | 46 (51%) | 22 (24%) | 12 (13%) |

Note. EPT: English Placement Test, APT: Arabic Placement Test

Table 2

Distribution of Person-fit Statistics

| Test | Final no. of items | Mean | SD | KS | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| EPT | 96 | .17 | .99 | .03 | -.30 | .08 |
| APT | 78 | .02 | .92 | .02 | -.16 | .09 |

Note. EPT English Placemat Test, APT: Arabic Placement Test,
KS : Kolmogorove-Smirnov Test of Normality

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)

TM034930

## I. DOCUMENT IDENTIFICATION:

Title: The Use of Person-Fit Statistics to Analyze Placement Tests

Author(s): Hamzeh Dodeen

Corporate Source: Paper presented at AERA - 2003 Chicago, IL.

Publication Date:

## 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 | The sample sticker shown below will be affixed to all Level 2A documents | The sample sticker shown below will be affixed to all Level 2B documents |
|---|---|---|
| PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | 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) | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) |
| 1 | 2A | 2B |
| Level 1 ☑ | Level 2A ☐ | Level 2B ☐ |
| Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy. | Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only | 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.*


| | | | |
|---|---|---|---|
| Signature: Hamzeh Dodeen | | Printed Name/Position/Title: Dr. Hamzeh Dodeen / Assistant Professor. | |
| Organization/Address: United Arab Emirates University (UAEU) UAE, Al-Ain, P.O. Box 15551 ~~Office of DVCAA~~ | | Telephone: 0097137042284 | FAX: 0097137641110 |
| | | E-Mail Address: hdodeen@hotmail.com | Date: 8-4-2003 |

Dr. Hamzeh Dodeen


ERIC Full Text Provided by ERIC

## 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, College Park**
**1129 Shriver Lab**
**College Park, MD 20742**



ERIC
Full Text Provided by ERIC