



March 1986

Assessing Child Maltreatment: The Role of Testing

Joel S. Milner
Western Carolina University

Follow this and additional works at: <https://scholarworks.wmich.edu/jssw>



Part of the [Social Welfare Commons](#), and the [Social Work Commons](#)

Recommended Citation

Milner, Joel S. (1986) "Assessing Child Maltreatment: The Role of Testing," *The Journal of Sociology & Social Welfare*: Vol. 13 : Iss. 1 , Article 6.

Available at: <https://scholarworks.wmich.edu/jssw/vol13/iss1/6>

This Article is brought to you for free and open access by the Social Work at ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.



Assessing Child Maltreatment: The Role of Testing

Joel S. Milner
Western Carolina University
Callowhee, North Carolina

ABSTRACT

Due to the recent development of test instruments designed to assist professionals in the evaluation of child maltreatment cases, social service professionals must become familiar with issues related to test construction and use. The purpose of the present paper is to provide the reader with a discussion of issues related to test selection and use. This article, however, is not intended to be a substitute for a basic understanding of the test validity and reliability. The paper begins with a review of different prevention modes and discusses how each mode uses test data. Next, test use as it relates to assessment is outlined. Within the context of assessment, types of test classification errors are discussed. Finally, a number of general test issues that can affect test results are presented.

Assessing Child Maltreatment: The Role of Testing

In the past decade, a large number of checklists, surveys, and test instruments have been developed to assist the protective service worker in the assessment and treatment of parents suspected of child abuse and neglect. While most of these instruments have little or no information on their validity and reliability, an increasing number of scales do provide appropriate psychometric data, which allows the test user to evaluate the usefulness and accuracy of the test. Several assessment tools on which some psychometric data has been accumulated include the Michigan Screening Profile of Parenting (MSPP) scale (Helfer et al., 1978; Schneider, 1982), the Conflict Tactics (CT) scales (Straus, 1979), the Childhood Level of Living (CLL) scale (Polansky et al., 1972; Polansky et al., 1978), the Parenting Stress Index (PSI) (Abidin, 1983), and the Child Abuse Potential (CAP) Inventory (Milner, 1980; Milner et al., 1984). These instruments measure behaviors extending from general problems in parenting (e.g., MSPP) to specific maltreatment problems, such as physical child abuse (e.g., CAP Inventory). Since these and

similar instruments will be available in increasing numbers in the future, the purpose of this article is to discuss the appropriate role such instruments should play in the assessment of child maltreatment cases. It is not the purpose of this paper to provide an extensive discussion of technical issues related to test validity and reliability. Rather, the focus of this paper will be on theoretical issues which will enable the reader to select and use tests appropriately once test validity and reliability has been determined.

Since the degree of emphasis on test use and the type of classification error a professional wishes to avoid varies as a function of the prevention mode, the present paper will begin with an exploration of the role of testing in primary, secondary, and tertiary prevention. Following this discussion, the role of testing in assessment will be presented. The assessment section, which will include a general paradigm for assessment, is provided to give the reader a conceptual guide to the appropriate use of test instruments in screening and diagnosis. Within the context of assessment, the different types of test misclassifications and their associated problems will be delineated. This information should enable the professional to deal more appropriately with classification errors. The article will conclude with a discussion of additional issues related to the selection and use of tests in the assessment of child maltreatment cases.

Types of Prevention

In the field of prevention, three types of prevention efforts have been distinguished. As previously mentioned, the three types are primary prevention, secondary prevention, and tertiary prevention.

Primary prevention assumes that all families in society are more or less at risk of abuse, neglect and/or problems in parenting because of our mobile, impersonal, and generally stressful society. Given this assumption, primary prevention is not concerned with screening or diagnostic activities. It is not concerned with the testing of abusive or neglectful individuals or with the preselection of at-risk groups suffering from poor parenting or other identifiable dysfunctions. Since all families are believed to be at risk, professionals involved in primary prevention are not concerned with misclassification issues. Rather, they are concerned with increasing the number of community support systems available to all families by promoting related legislation and resource

allocation. Advocates of primary prevention stress broad educational and social policy interventions designed to lessen the impact of social and psychological stresses on all families. Their range of interests, therefore, extend from concern about the economic resources available to families so that adequate shelter, food, and clothing are available, to concerns about parent education programs in the schools, to maintaining public awareness of child maltreatment, and to considerations of cultural differences that may affect the quality of family life.

Secondary prevention, in contrast, does not assume that all families are to some degree at risk of abuse, neglect, and/or problems in parenting. Instead, it assumes that only certain families are seriously at risk of child maltreatment and that these families can and should be identified and offered services. Secondary prevention strategies, which are received voluntarily, include family life education programs, counseling, self-help groups, home health visitors, crisis day care, etc. In an attempt to provide services to those who are at risk, advocates of secondary prevention are often interested in using some form of screening criteria or testing to identify groups of at-risk individuals or families. In the screening process, there is less concern about false-positive classifications; that is, identifying an individual as at risk who actually is not. Instead, assessment criteria tend to be overinclusive and test cut-off scores are set low so that most at-risk subjects are selected. This approach produces few false-negative classifications; that is, identifying an individual as not at risk who actually is at risk. The goal is to construct an at-risk group without eliminating any at-risk individuals and to offer the at-risk group direct intervention services without applying labels and without diagnosis. Intervention services may be offered to the child, to the parents, or to the total family system in order to prevent the occurrence of child maltreatment.

In tertiary or legal prevention, society attempts through legislation to prevent child abuse and neglect and to prevent its reoccurrence. Legislative restrictions (i.e., abuse and neglect reporting laws) are placed on the caretaker-child relationship, and if abuse and neglect is believed to have occurred, then action supposedly follows. That is, after an abusive and/or neglectful act has occurred a report is made, an assessment leading to a diagnosis is completed, and intervention and/or adjudication follows. During this legal process, correct diagnosis with the elimination of false-positive classifications is a primary goal. However, as one type of error decreases (i.e., false positives) the

alternate error (i.e., false negatives) will increase, which in this case means that there will be an increase in the number of abusive and neglectful parents that go undetected. This error is permitted because of the often severe consequences of adjudication. For example, following adjudication for maltreatment, children may be removed, parents may go to jail, and parents may be required to submit to intervention (e.g., therapy). Thus, while avoidance of false-positive classifications provides more protection for the adults involved, the cost is an increase in the number of children who will not be protected from actual abusive and neglectful caretakers who go undetected. The problems generated by this dilemma will likely remain an issue of heated debate for decades to come.

As can be seen from the foregoing discussion, the importance of assessment methods and associated misclassification errors varies with the type of prevention efforts employed. To better understand the appropriate use of screening and diagnostic testing in each prevention mode, screening and diagnosis will be defined and discussed within the context of an assessment paradigm. This information is provided to give the reader a conceptual understanding of problems inherent in screening and diagnosis.

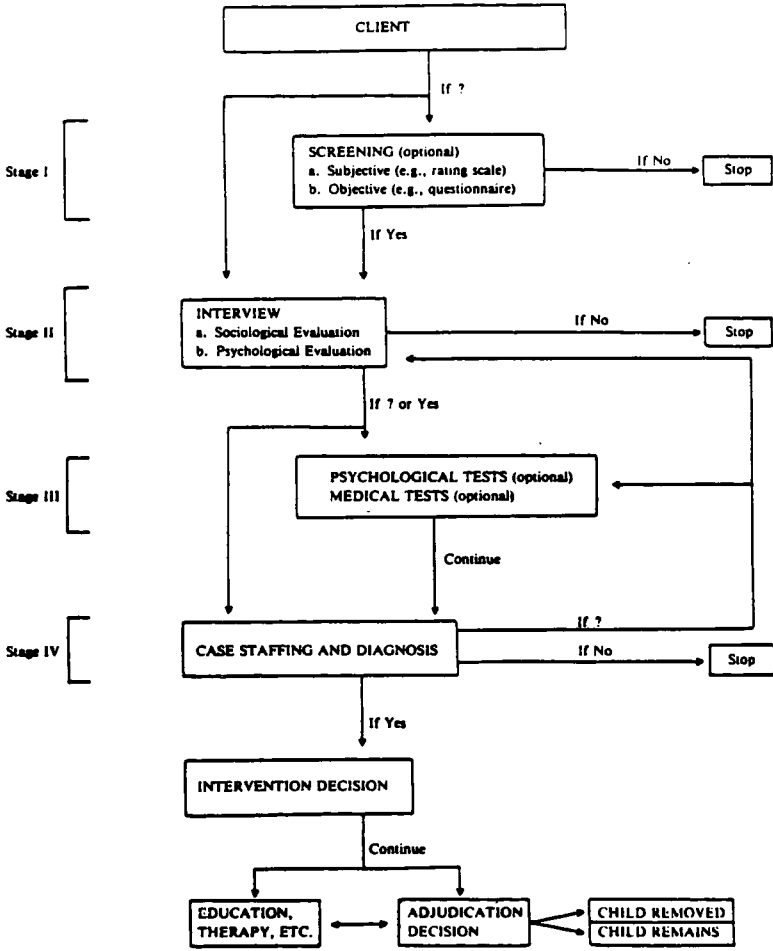
Stages of Assessment

In assessment, screening is a term employed to describe a rapid, often rough selection process (Anastasi, 1982). Usually, screening refers to a preliminary attempt to determine if a personality characteristic or behavior is present or absent in a given individual or group. In most cases, the screening activity occurs as the initial stage of an assessment process.

Diagnosis is another term describing an activity which is part of the assessment process. The term diagnosis, however, is employed to designate a more intensive and comprehensive evaluation process which occurs in the last stage of assessment.

A paradigm describing the typical stages of assessment which begins with screening (Stage I) and leads to diagnosis (Stage IV) is outlined in Figure 1. This assessment approach provides a strategy for the clinical practitioner to follow. As with any decision process, the strategy defines what the practitioner will do in any number of possible situations.

Figure 1 A Paradigm Describing Typical Assessment Stages Leading to Diagnosis and Outcome



Insert Figure 1 here

While the assessment process has been greatly simplified in Figure 1, inspection of the flow chart reveals that this assessment paradigm approaches decisions sequentially. Until the end of the process is reached, no commitment to an individual diagnosis is made. If after one stage of information gathering the individual is moved to another stage, new information about the client is usually gained. The addition of new data may and often does modify the ultimate diagnosis. Even at the last stage of assessment, if the information gathered remains ambiguous or incomplete, the strategy provides for a return to a previous stage in order to obtain needed data. For example, if during the case staffing/diagnosis stage (Stage IV), it is determined that additional history, personality testing, and/or medical tests are needed, the process can return to the appropriate prior assessment stage. Flexibility and completeness are the key features of this approach so that when a final case diagnosis is made, misclassifications are minimized.

While reviewing this assessment paradigm, it is important to understand the distinction between screening and diagnosis. Screening is the initial process (Stage I) and employs a technique which places individuals in one of two categories. That is, either a person is believed to have a specified characteristic (e.g., risk for abuse) or he does not. No other outcome is provided. Thus, placement is dichotomous. While a variety of screening approaches have been developed (e.g., rating scales, tests, etc.), most screening procedures are designed to be brief in terms of items and quickly administered in contrast to more comprehensive and time consuming diagnostic procedures.

During the process of screening individuals into one of the dichotomous categories, two correct and two incorrect classifications are possible. Figure 2 describes the four screening outcomes that can occur. The screening procedures may correctly identify an

Insert Figure 2 here

at-risk client as being at risk of abuse (outcome A) or the screening may misclassify the at-risk individual as not at risk (outcome C). The ability to correctly identify an at-risk individual is known as the sensitivity of the screening procedure. The misclassification or error of designating

Figure 2 Four Possible Screening Outcomes

		Actual	
		Client At Risk	Client Not At Risk
Screened	Client At Risk	A	B
	Client not At Risk	C	D

- A Correct classification of at-risk status (i.e., sensitivity)
- B Misclassification of at-risk status (i.e., false-positive classification)
- C Misclassification of at-risk status (i.e., false-negative classification)
- D Correct classification of non-risk status (i.e., specificity)

the at-risk individual as not at risk is called a false-negative classification. Other screening outcomes involve correctly identifying a non-at-risk client as not at risk (outcome D) or misclassifying the non-at-risk as at risk (outcome B). The ability to correctly identify a non-at-risk individual is known as the specificity of the screening procedure. The misclassification or error of calling the non-at-risk individual as at risk is called a false-positive classification.

The type of error allowable varies from stage to stage in the assessment process. False-negative classifications are usually avoided during the screening stage, while false-positive classifications are usually considered undesirable during the diagnosis stage.

During screening (Stage I), false-positive classifications are generally allowed and false-negative classifications avoided because at this stage the goal is to avoid missing any actual cases while reducing the often large pool of clients who will continue to the next stage of assessment. Since it is believed that subsequent stages will eliminate any false-positive classifications, the primary concern is that the rating scale criteria or test cutting score be set low so that the screening procedure will avoid missing actual at-risk cases.

In contrast, during the diagnostic stage (Stage IV) the focus changes. It is now important to avoid false-positive classifications because there will not be any subsequent stages to check on questionable cases. Thus, during the diagnostic stage, the criteria for inclusion in the maltreatment group is expected to be more comprehensive and complete. Again, at the diagnostic stage, the goal is to avoid false-positive classifications and the accompanying false accusation of an innocent client.

When classification errors are discussed, the type of error allowed is also affected by the type of intervention that follows the positive classification (e.g., the parent is an abuser or is at risk). If the intervention that follows involves the offer of such things as advice, education, counseling, therapy, child care and/or referral to community resources and services (as is the case in secondary prevention), false-positive selections may not be considered very damaging when weighted against the possible preventive benefit to children who may otherwise be mistreated. Further, when resources are limited and diagnosis is not a goal, a screening procedure, even one with only moderate predictive validity, is beneficial since it allows professionals to

select from a large population a subgroup of individuals who are most likely to benefit from the use of available resources. In this case, an increase in the cost/benefit ratio of the program would be expected as an ancillary outcome because those most in need of services are targeted. On the other hand, the use of more intrusive interventions involving labeling and reporting with department of social services and court involvement, which occurs in tertiary prevention, makes the elimination of false positives a necessary and important goal. With more intrusive intervention, the false-positive classification can have severe negative consequences and harmful effects on the clients and families involved. Thus, the type of classification error the practitioner is most concerned about will vary as a function of the stage of assessment and the type of intervention (i.e., secondary or tertiary prevention) that follows assessment.

General Guides to Test Use

As previously noted, the information in the present article is meant to be supplemental to basic understanding of test validity and reliability and is not a substitute for such knowledge. For a detailed guide to the understanding and use of psychological tests, the reader is referred to the Standards for Educational and Psychological Testing (American Psychological Association, 1985). Included in this document are discussions of technical standards of test construction and professional standards for test use. However, there are several general considerations in the selection of tests that need to be discussed and that are within the scope of the present paper.

Individual Versus Group Classifications

When test instruments are selected, the professional must be cognizant of the differences between the ability of a test to produce group differences and the ability of a test to successfully classify individuals. Many psychological tests, some of which have little ability to screen individual clients, can produce highly significant group differences. For example, it is possible for a given test to produce mean scale scores for abuse and control groups which are significantly different, while correctly classifying less than 50% of the actual abusers as abusive. The professional must have available the information on individual misclassification rates, as presented in Figure 2, in order to properly evaluate a test. If such information is not available in the test manual, a

test should not be utilized for classifying individuals.

What Behavior (Construct) Does the Test Measure

When selecting a test instrument, it is important for the professional to know what attitudes and behaviors the test is measuring. For example, a study may indicate that a test successfully screened individuals that were abusive and nonabusive in a given sample. The test, however, may be measuring constructs such as distress in order to predict group membership. In this example, it would be necessary for the test author to demonstrate that the distress measured relates primarily to abusive behavior and not to a general measure of distress. If general distress is measured, when a larger sample of individuals is screened, clients with personal distress (e.g., a death in the family, personal injury/illness, etc.) that is not necessarily related to child abuse potential would obtain elevated scores and be misclassified as abusive.

Use of Test Scores in Diagnosis

The professional must be aware that a test score alone should never be employed to make a diagnosis. An accumulation of data obtained from multiple sources, as described in Figure 1, must be used. This is especially important when a case is adjudicated. In some instances when assessment data remain ambiguous and a decision must be made to remove a child from the home (or to return a child), there may be pressure to employ a raw test score as a basis for the decision. This is a particularly inappropriate action given the previous discussion of false-positive and false-negative classifications that exist even with the psychometrically "best" tests. Further, each individual test score is only an approximation of an individual's "true" score and is known to contain measurement error. That is, if the test is administered again to the same individual, the test score will likely be higher or lower. Test classification errors and individual test measurement errors are of particular concern when cut-off scores are employed and individuals earn scores that are close (either just above or just below) to the cut-off score. Even when this is not the case, classification and measurement errors make the use of individual raw scores inappropriate for decision making.

What is the Base rate of the Sample Tested

A major technical problem that often is not understood is the effect of

baserates on test misclassifications. Baserate refers to how often a given behavior occurs in a specific sample. When considering the appropriateness of testing, it is not sufficient to merely consider the prevention mode and test misclassification rates. The prevalence of the target behavior (e.g., maltreatment) in the sample tested must also be considered. Simply stated, test instruments are most useful when the sample tested contains approximately 50% of the deviant subjects. In a social service setting where about one half of those reported for maltreatment are confirmed, testing would be appropriate, at least in terms of baserates. However, use of the same test instrument in the general population might not be warranted.

For example, if the baserate for child maltreatment in the general population is 5% and a test misclassifies 20% of both the maltreating and normal parents, the error rate in the group selected as maltreating in the general population will be much higher than 20%. This is true because in a sample of 100 parents, where the test has a 20% misclassification rate for normals, approximately 19 (20% of 95 normal parents) of the individuals that are not maltreating will be selected as maltreating parents. Likewise with a 20% error rate for maltreatment group selection, only 4 of the 5 (5% baserate) maltreating parents will be selected. Thus, for the general population where the baserate for maltreatment is a hypothetical 5% and the test misclassification rate is 20%, the test will select about 23 individuals (19 normal and 4 maltreating parents) as maltreating parents and will be correct in only 4 of the 23 cases selected. Even though it is true that only one of the subjects screened as normal was misclassified, incorrectly classifying 19 out of 23 parents, 19 of which are normal, as maltreating is unacceptable. While multi-stage screening, which will not be discussed in this paper, provides a partial solution for low baserate problems, it is apparent that the error rate for a test can vary dramatically when populations with different baserates are tested.

Other Test Considerations

There remain several other issues that should be considered in the selection of a test instrument. The readability level of the instrument must be acceptable. The test should contain validity subscales (e.g., faking good, faking bad, and random response scales). Validity subscales, especially those that measure faking behavior, are important in settings where there is a likelihood that the client will be motivated to distort

his/her answers in a socially desirable manner. In social service settings, the client may be particularly guarded in his/her responding if there is an investigation for child maltreatment. Thus, misclassification rates, especially false negatives, can be expected to increase if a test does not contain validity checks for response distortions. Finally, the selected test should be relatively free from bias due to demographic variables (e.g., gender, age, educational level, ethnic background, etc.). To the extent that special populations have different test scores, appropriate norm scores should be provided in the test manual.

Conclusion

Recent increases in the number of screening and test instruments available to social service professionals involved in secondary and tertiary prevention of child maltreatment is an advancement which has great potential for use and misuse. The burden for appropriate use is placed squarely on the professionals involved. If these new tools are misused, they can result in extensive damage to the families served. In contrast, if these tests are employed by professionals who have an adequate knowledge of test construction and test use, test results can be a valuable source of information that may be combined with other data to make more accurate assessment and treatment decisions.

REFERENCES

- Abidin, Richard R. Parenting Stress Index Manual, Charlottesville, Virginia: Pediatric Psychology Press, 1983.
- American Psychological Association, Standards for Educational and Psychological Testing. Washington, D. C.: American Psychological Association, 1985.
- Anastasi, Anne. Psychological Testing, New York: Macmillan Publishing Company, 1982.
- Helfer, Ray E., James K. Hoffmeister, and Carol Schneider, MSPP: A Manual for use of the Michigan Screening Profile of Parenting. Boulder, Colorado: Test Analysis and Development Corporation, 1978.
- Milner, Joel S. The Child Abuse Potential Inventory: Manual. Webster,

NC: Psytec Corporation, 1980.

Milner, Joel S., Ruth G. Gold, Catherine Ayoub, and Marion M. Jacewitz. "Predictive Validity of the Child Abuse Potential Inventory." Journal of Consulting and Clinical Psychology 52: 879-884, 1984.

Polansky, Norman A., Robert D. Borgman, and Christine Desaix, Roots of Futility. San Francisco: Jossey-Bass, 1972.

Polansky, Norman A., Mary A. Chalmers, Elizabeth Bittenwieser, and David Williams. "Assessing the Adequacy of Child Caring: An Urban Scale." Child Welfare 57: 439-449, 1978.

Schneider, Carol J. "The Michigan Screening Profile of Parenting." pp. 157-174 in R. Starr (Ed.), Child Abuse Prediction. Cambridge, Massachusetts: Ballinger Publishing Company, 1982.

Straus, Murray A. "Measuring Intrafamily Conflict and Violence: The Conflict Tactics (CT) Scales." Journal of Marriage and the Family 41: 75-88, 1979.

Footnotes

This project was supported in part by Grant No. 1R01 MH34252 awarded by the National Institute of Mental Health.

Request reprints from Joel S. Milner, Department of Psychology, Western Carolina University, Cullowhee, North Carolina 28723.