DOCUMENT RESUME

ED 246 129 TM 840 410

AUTHOR Fuchs, Douglas; And Others
TITLE Prediction of Suboptimal Test Performance among Handicapped Children: An Exploratory Investigation.
PUB DATE Apr 84
NOTE 19p.; Portions of this paper were presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Education; *Examiners; Experimenter Characteristics; Language Handicaps; *Performance Factors; Prediction; Preschool Education; *Rapport; Speech Handicaps; Testing; *Test Results
IDENTIFIERS *Examiner Familiarity

ABSTRACT

This study employed a multiple regression to predict examinees' differential performance when tested by familiar and unfamiliar examiners. Subjects were 32 preschool and school-age handicapped children, each of whom had been tested on the Clinical Evaluation of Language Functions, once by a familiar and once by an unfamiliar tester, within a crossover design. Teachers' ratings, subjects' self-reports, an anxiety scale, and a sociometric procedure accounted for nearly 40 percent of the variance in differential test performance. Implications for valid assessment of handicapped children are discussed. (Author)

Prediction of Suboptimal Test Performance among Handicapped Children: An Exploratory Investigation

Douglas Fuchs

Clark University

Lynn S. Fuchs

Wheelock College

Marilyn L. Blaisdell

Becker Junior College

Portions of this paper were presented at the annual meeting of the American Educational Research Association, New Orleans, April 1984.

We are grateful to the many people involved in the execution of this study, particularly to Patti Goodman, Lori Lefleur, Dana Siegel, and Nina Stolzenberg, who collected much of the data, and to Dick Gilmore and Ann Dailey, who helped identify subjects and contact teachers and parents.

Requests for reprints should be sent to Douglas Fuchs, Department of Education, Clark University, 950 Main Street, Worcester, MA 01610.

Running head: Prediction of Suboptimal Performance

Abstract

This study employed a multiple regression to predict examinees' differential performance when tested by familiar and unfamiliar examiners. Subjects were preschool and school-age handicapped children, each of whom had been tested on the Clinical Evaluation of Language Functions, once by a familiar and once by an unfamiliar tester, within a crossover design. Teachers' ratings, subjects' self-reports, an anxiety scale, and a sociometric procedure accounted for nearly 40% of the variance in differential test performance. Implications for valid assessment of handicapped children are discussed.

Prediction of Suboptimal Test Performance among Handicapped Children: An Exploratory Investigation

There is considerable evidence (e.g., Fuchs, Featherstone, Garwick, & Fuchs, 1984; Fuchs, Fuchs, Garwick, & Featherstone, 1983; Olswang & Carpenter, 1978; Stoneman & Gibson, 1978) that the unfamiliarity of an examiner negatively affects handicapped children's optimal absolute performance. Moreover, a recent investigation (Fuchs, Fuchs, Power, & Dailey, 1983) demonstrated that tester unfamiliarity depresses handicapped, but not nonhandicapped, children's test performance, thereby indicating an examiner's unfamiliarity constitutes a negatively biasing condition and threatens the validity of handicapped students' test performance.

This evidence supports the view that examiners be familiar with handicapped children prior to testing. Whereas testers might view this prescription as conceptually sound, it also is likely many would perceive it as infeasible since examiners frequently operate within severe time constraints (see AERA, APA, & NCME, 1983, p. 14-2). However, if it were possible to predict which handicapped students are likely to perform suboptimally with an unfamiliar examiner relative to performance with a familiar examiner, then testers might establish pretest contact with only a subgroup of these pupils.

The purpose of this investigation was to begin to develop a basis for such a prediction. Previous predictive efforts have been sporadic and inconclusive (cf. Feldman & Sullivan, 1971; Fuchs, Fuchs, Garwick, & Featherstone, 1983; Fuchs, Featherstone, Garwick, & Fuchs, 1981). Also, they typically have included a teacher rating as the only prediction measure. In contrast, the present investigation sought to predict children's differential test performance on the basis of information (a) gathered from multiple informants (i.e., parents, peers, subjects, and teachers) and (b) generated by qualitatively different

procedures (i.e., parent ratings, peer nominations, self-reports, and teacher ratings).

Method

Subjects

Subjects were 32 handicapped children from a larger study (Fuchs et al., 1983), in which 64 subjects comprised four groups: 16 handicapped preschoolers (HP), 16 handicapped school-age children (HS), 16 nonhandicapped preschoolers (NP), and 16 nonhandicapped school-age children (NS). The HPs and HSs were moderately to profoundly speech and/or language impaired, performed within the normal range on individually administered IQ tests, and were participating in language programs (preschool or elementary school levels) in the same public educational collaborative. NPs and NSs were drawn from a large college-affiliated nursery school and public elementary school, respectively. All subjects were Caucasian, English-speaking, and from predominantly middle-class families located in 5 contiguous towns in Central Massachusetts. The mean CAs for handicapped and nonhandicapped subjects were 77.44 (SD = 24.91) and 76.91 (SD = 24.49) months, respectively. Average CAs for preschool and school-age subjects were 57.16 (SD = 7.20) and 97.19 (SD = 18.48) months, respectively. Whereas a two-way analysis of variance (ANOVA; handicapped vs. nonhandicapped and preschool vs. school-age) revealed a significant difference between the CAs of preschool and school-age subjects, F (1,60) = 128.28, $p < .001$, there was no significant disparity in CA between handicapped and nonhandicapped children, $F(1,60) = .02$, ns, and no significant interaction $F(1,60) = .00$, ns. Additionally, identical numbers of male (N = 10) and female (N = 6) subjects constituted the four subject groups.

As part of this larger study, each handicapped and nonhandicapped subject had been tested by 2 of 32 speech clinicians on the Clinical Evaluation of Language Functions. Each subject was tested twice, once by a familiar examiner and once by an unfamiliar tester, within a crossover design. Also, the order in which examiners tested familiar and unfamiliar subjects and the time of day subjects were assessed were controlled. Findings of this larger study indicated that, in comparison to nonhandicapped children, handicapped subjects performed significantly stronger when tested by examiners with whom they were familiar. There was no significant interaction between familiarity and CA or among familiarity, handicapping condition, and CA.

Measures

The Clinical Evaluation of Language Functions (CELF; Semel-Mintz & Wiig, 1982) is a comprehensive language test comprised of two scales, processing (PS; i.e., auditory comprehension) and production (PC; i.e., verbal expression). Semel-Mintz and Wiig (1982) reported internal consistency of .82 and a test-retest reliability coefficient of .96. Pilot administrations of the CELF were conducted with language-impaired preschoolers (CAs: 3-10 to 4-8) to determine the possibility of a "floor" effect. Results indicated the CELF was appropriately difficult for the study's sample. Examiners were trained to administer the CELF in one 3-hour training session, which was conducted by a certified speech clinician.

Iowa Social Competency Scales: Maternal Preschool and School-age Forms (Iowa; Pease, Clark, & Crase, 1982) assesses children's compliance in social situations with parents, siblings, peers, and unfamiliar persons. Specifically, the preschool form explores whether children are socially activating, hypersensitive, or adaptive in unfamiliar situations and/or with unfamiliar

persons; capable of sharing; and cooperative. The school-age form examines children's behavior with respect to task orientation, egocentricity, leadership, physical activity, affection toward parents, and apprehension. The frequency with which these behaviors are displayed is rated by mothers on a scale of 1 (infrequent) to 5 (always). For the Preschool Form, Pease et al. (1982) report median Spearman-Brown reliability estimates for total and unique variances to be .78 and .75, respectively. Pease et al. obtained similar reliability estimates for the School-age Form: .75 and .74 for total and unique variances, respectively.

Perceived Competence Scale for Children (PCSC; Harter, 1982) is a self-report measure developed for children between 8 and 15 years. It explores respondents' self-esteem (PCSC-E) and perceived competence in the cognitive (PCSC-C), social (PCSC-S), and physical (PCSC-P) domains. Harter (1982) reports Cronbach's alpha of .73, .76, .78, and .83 for each of these subscales, respectively. The Pictorial Scale of Perceived Competence and Acceptance for Young Children (Harter & Pike, 1981) comprises two additional forms: "P-K" for ages 4 and 5, and "1-2" for 6 and 7 year olds. These focus on areas of competence very similar to those constituting the PCSC. Whereas children respond to written statements on the PCSC, the latter two forms employ a pictorial format. Harter and Pike (1981) report overall Cronbach's alpha of .88 and .87 for the "P-K" and "1-2" forms, respectively. No distinction will be made below between Harter's parallel scales for younger and older children; all will be identified as "PCSC."

Teacher Rating Scale of Children's Actual Competence (TRS; Harter, 1982) parallels the PCSC in substance. The present study employed the cognitive (TRS-C), social (TRS-S), and physical (TRS-P) subscales. On the TRS, teachers

rate children on a scale of 1 (least competent) to 4 (most competent). In this investigation, the handicapped children's special education teachers completed the TRS. Harter (1982) reports internal consistency (KR20) coefficients of .96, .93, and .94 for the TRS-C, TRS-S, and TRS-P subscales, respectively.

Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1978) comprises 28 anxiety-related questions, such as "It's hard for me to get to sleep at night," "I feel sick to my stomach a lot," and "A lot of people don't like me." The questions are read to the children and are scored either "Yes" or "No" by the examiner. An internal consistency (KR20) coefficient of .83 was reported (Reynolds & Richmond, 1978).

Sociometric data were collected by means of a nominative approach that was adapted from Hymel and Asher (1977). Preschool and kindergarten subjects and their handicapped classmates selected three peers with whom they most liked to play, or Positive Nomination (PNOM), and another three students with whom they least liked to play, or Negative Nomination (NNOM). School-age subjects and their handicapped classmates nominated three children from their classroom for the following six categories: Like (and don't like) to work with in class; like (and don't like) to play with on the playground; and like (and don't like) to invite home after school. Preschool and school-age subjects' PNOM and NNOM scores were the total numbers of nominations they received from their classmates. Asher, Singleton, Tinsley, and Hymel (1979) and Oden and Asher (1977) provide various documentation for the stability of these nominative procedures.

Procedure

Data collectors (DCs) were four female graduate and undergraduate students. They participated in five training sessions (14 hours), during which

they were trained to administer various instruments and procedures. One week after subjects were tested on the Clinical Evaluation of Language Functions (CELF), as part of the larger investigation, DCs administered the PCSC. PNOM and NNOM procedures were conducted in small groups or individually, depending upon children's reading skills. At both the preschool and school-age special education sites, all of the subjects' classmates participated in the nominative procedures. The TRS was distributed to the teacher with whom each subject spent the most time. The Iowa was completed by subjects' mothers or maternal guardians. The RCMAS was administered to subjects by their familiar speech clinician at the end of the administration of the CELF in the earlier phase of the investigation.

Results

The dependent variable of interest, the difference between children's performance in the familiar and unfamiliar test conditions, was converted to a regressed-adjusted residualized variable to avoid problems inherent in working with difference scores (Cohen & Cohen, 1975). In this process, scores in the unfamiliar condition, the covariate set (A), were regressed on scores in the familiar condition, the dependent variable (Y), resulting in a new set of familiar condition scores. This new estimate then was subtracted from the original familiar condition score, creating a regressed-adjusted familiarity score (RAFS), which was entered into the analyses presented below. This process variously is referred to as "residualization" of Y by A, "adjusting" Y for A, and Y "with A held constant statistically."

Means and standard deviations are displayed in Table 1 for RAFS and the following variables: (a) PCSC-C, PCSC-P, PCSC-S, and PCSC-E subscales of the Perceived Competence Scale for Children and the Pictorial Scale of Perceived

Competency and Acceptance for Young Children ("P-K" and "1-2"); (b) TRS-C, TRS-P, and TRS-S subscales of the Teacher Rating Scale of Children's Actual Competence; (c) negative (NNOM) and positive (PNOM) nominations; (d) Revised Children's Manifest Anxiety Scale (RCMAS); and (e) the Iowa Social Competency Scales (Iowa) for preschool and school-age children.

Insert Table 1 about here

Correlational analyses were conducted to examine the relations among RAFS and the 11 psychosocial variables. These Pearson product-moment correlations are shown in Table 2. Five variables correlated with RAFS at a value of at least $\pm.20$: $r(32) = .50$, $p = .002$ for TRS-C; $r(32) = .29$, $p = .055$ for TRS-P; $r(32) = .26$, $p = .073$ for NNOM; $r(32) = -.33$, $p = .034$ for CMAS; and $r(32) = -.20$, $p = .137$ for Iowa.

Insert Table 2 about here

Of these five variables, four were entered into a forward stepwise multiple regression to determine what portion of the total variance in RAFS was accounted for by these predictors. TRS-P, the fifth variable, was excluded because of its high intercorrelations with three of the other variables entered into the multiple regression (with TRS-C, $r(32) = .38$, $p = .016$; with NNOM, $r(32) = -.25$, $p = .085$; with Iowa, $r(32) = -.36$, $p = .021$). Results of the multiple regression are shown in Table 3. TRS-C accounted for 25% of the total variance, followed by CMAS (6%), Iowa (5%), and NNOM (3%). The effects due to these four variables were significant

(see Table 3). Together, the four predictors accounted for 39% of the total variance in RAFS.

Insert Table 3 about here

Discussion

Findings indicate moderate relations between children's differential performance across familiar and unfamiliar examiner conditions and the following variables: the Cognitive subscale (TRS-C) of the Teacher Rating Scale of Children's Actual Competence, Negative Nomination (NNOM), Revised Children's Manifest Anxiety Scale (RCMAS), and the Iowa Social Competency Scales (Iowa).

This suggests several distinctive characteristics of speech-and/or language-handicapped children whose differential performance in favor of familiar examiners reflects suboptimal performance with unfamiliar examiners. First, they seem to be comparatively competent pupils, as indicated by the strong positive correlation between TRS-C and the regressed-adjusted familiarity score (RAFS). Second, they appear to be relatively free of chronic anxiety, as evidenced by the inverse relation between RCMAS and RAFS. Last, they seem to have poor social skills, as suggested by the high number of negative nominations they received from classmates and low ratings they were assigned by their mothers. Moreover, the negative correlation between RAFS and subjects' ratings of themselves on the social subscale of the Perceived Competency Scale for Children suggests they are aware of their poor social skills. Taken together, these results suggest a portrait of an examinee who, although not skilled in social relations with peers and adults, has the capacity to recognize the familiar examiner as accepting and supportive

and, perhaps because of relatively low anxiety, can respond adaptively in the familiar condition.

The important role of cognitive competence in this portrait is corroborated by Mischel (1973), who describes certain individuals' cross-situational variability as a highly refined "discriminative facility." Its apparent importance also is consonant with results from an investigation conducted by Harter (1967), who employed retarded, normal, and bright children and found the normal and bright children learned more rapidly in a social condition, whereas the retarded children learned more quickly in a non-social condition.

The four measures involved in this portrait, TRS-C, NNOM, RCMAS, and Iowa, accounted for nearly 40% of the variance in differential performance with familiar and unfamiliar examiners. This result compares favorably with previous investigations (e.g., Fuchs, Fuchs, Garwick, & Featherstone, 1981) and should encourage continued work on the development of an instrument that would predict, among select groups, children whose test performance may be improved through familiarization with an examiner.

Nevertheless, statistical and conceptual issues require words of caution. When multiple regression is used with a relatively small sample, the regression coefficients tend to be unstable from one sample to another (Kerlinger & Pedhazur, 1973). While the present use of multiple regression seems justified within the context of this exploratory study, replication with a larger sample is necessary.

More theoretically, this study identified suboptimal performers by their stronger performance with familiar testers. This is a limited conceptualization of suboptimal performance because it precludes recognition of examinees who may perform suboptimally in <u>both</u> familiar and unfamiliar examiner condi-

tions. Further research on explicating and predicting suboptimal test performance might employ alternate and more comprehensive definitions of the construct.

Second, the orientation of this investigation implies that speech-and/or language-impaired children performed more strongly with familiar testers because of factors originating within themselves. However, recent evidence indicates examiners' attitudes and behaviors sometimes affect children's test performance (cf. Sattler, 1974). This bears on the nature of the above portrait of examinees performing more strongly with familiar testers. Rather than view differential performance as an expression of examinees' capacity to respond appropriately to subtle changes in situational contingencies, differential performance with familiar examiners may relate to testers' selective, facilitating behavior.

Specifically, testers familiar with the more cognitively competent handicapped child may provide more encouragement (e.g., Thomas, Hertzig, Dryman, & Fernandez, 1971), permit greater amounts of response time (e.g., Fuchs, Zern, & Fuchs, 1983), or use less stringent scoring criteria (e.g., Fuchs & Fuchs, in press; Sattler & Winget, 1970). Future research should address the challenging task of exploring simultaneously both examinee characteristics and examiner behaviors to understand better how to optimize handicapped children's test performance.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1983, February). Draft: Joint technical standards for educational and psychological testing. (Available from APA, Office for Scientific Affairs, 1200 17th St., N.W., Washington, D.C. 20036.)

Asher, S.R., Singleton, L.C., Tinsley, B.R., & Hymel, S. (1979). A reliable sociometric procedure for preschool children. Developmental Psychology, 15, 443-444.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Feldman, S.E., & Sullivan, D.S. (1971). Factors mediating the effects of enhanced rapport on children's performance. Journal of Consulting and Clinical Psychology, 36, 302.

Fuchs, D., Featherstone, N., Garwick, D.R., & Fuchs, L.S. (1984). Effects of examiner familiarity and task characteristics on speech and language-impaired children's test performance. Measurement and Evaluation in Guidance, 16, 198-204.

Fuchs, D., Featherstone, N., Garwick, D.R., & Fuchs, L.S. (1981). The importance of situational factors and task demands to handicapped children's test performance (Research Report No. 54). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Documentation Reproduction Service No. ED 203-587)

Fuchs, D., Fuchs, L.S., Garwick, D.R., & Featherstone, N. (1983). Test Performance of language-handicapped children with familiar and unfamiliar examiners. Journal of Psychology, 114, 37-46.

Fuchs, D., Fuchs, L.S., Power, M.H., & Dailey, A.M. (1983). Systematic bias in the assessment of handicapped children (Research Report No. 134). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Fuchs, D., Zern, D.S., & Fuchs, L.S. (1983). A microanalysis of participant behavior in familiar and unfamiliar test conditions. Diagnostique, 8, 159-169.

Fuchs, L.S., & Fuchs, D. (in press). Examiner accuracy during protocol completion. Journal of Psychoeducational Assessment.

Harter, S. (1982). The perceived competence scale for children. Child Development, 53, 87-97.

Harter, S. (1967). Mental age, IQ, and motivational factors in the discrimination learning set performance of normal and retarded children. Journal of Experimental Child Psychology, 5, 123-141.

Harter, S., & Pike, R. (1981). The pictorial scale of perceived competence and acceptance for young children: Manual. Denver: University of Denver.

Hymel, S., & Asher, S.R. (1977, March). Assessment and training of isolated children's social skills. Paper presented at the annual meeting of the Society for Research in Child Development, New Orleans.

Kerlinger, F.N., & Pedhazur, E.J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. Psychological Review, 80, 252-283.

Oden, S., & Asher, S.R. (1977). Coaching skills for friendship making. Child Development, 48, 495-506.

Ofswang, L.B., & Carpenter, R.L. (1978). Elicitor effects on the language obtained from young language-impaired children. Journal of Speech and Hearing Disorders, 43, 76-88.

Pease, D., Clark, S., & Crase, S.J. (1982). Iowa social competency scales: School-age and preschool manual. Ames, IA: Iowa State University Research Foundation.

Reynolds, C.R., & Richmond, B.O. (1978). What I think and feel: A revised measure of children's manifest anxiety. Journal of Abnormal Child Psychology, 6, 271-280.

Sattler, J.M. (1974). Assessment of children's intelligence. Philadelphia: W.B. Saunders.

Sattler, J.M., & Winget, B.M. (1970). Intelligence testing procedures as affected by expectancy and IQ. Journal of Clinical Psychology, 26, 446-448.

Semel-Mintz, E., & Wiig, E.H. (1982). Clinical evaluation of language functions. Columbus, OH: Charles E. Merrill.

Stoneman, Z., & Gibson, S. (1978). Situational influences on assessment performance. Exceptional Children, 46, 166-169.

Thomas, A., Hertzig, M.E., Dryman, I., & Fernandez, P. (1971). Examiner effect in IQ testing of Puerto Rican working-class children. American Journal of Orthopsychiatry, 41, 809-821.

Table 1

Means and Standard Deviations on Measures

Measure	Mean	Standard Deviation
RAFS	6.63	35.95
PCSC-C	20.31	3.13
PCSC-P	19.63	3.58
PCSC-S	19.00	4.26
PCSC-E	19.47	3.67
TRS-C	16.34	4.80
TRS-P	15.06	4.28
TRS-S	14.50	4.30
NNOM	15.06	10.75
PNOM	17.09	9.73
CMAS	15.25	6.92
Iowa	3.35	0.47

Table 2

Intercorrelations between Test Measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12
RAFS												
PCSC-C	.01											
PCSC-P	.10	.43										
PCSC-S	-.14	.20	.44									
PCSC-E	.09	.44	.50	.47								
TRS-C	.50	.05	.20	-.15	-.07							
TRS-P	.29	-.05	.19	-.11	-.08	.38						
TRS-S	.03	-.27	.33	.02	.06	.06	.49					
NNOM	.26	.03	-.15	-.08	.21	.13	-.25	-.12				
PNOM	-.04	-.18	.09	-.10	-.21	.17	.04	.14	-.06			
CMAS	-.33	-.03	.00	-.11	-.10	-.18	.09	.19	-.11	.19		
Iowa	-.20	-.09	-.16	.00	.03	-.02	-.36	-.14	.00	.08	-.20	

Table 3

Regression Analysis of Predictors on Regressed-Adjusted Familiarity Score

Source	Multiple R	R^2 cumulative	R^2 Change	B	β	F(1,30)	p<
TRS-C	.50	.25	.25	2.48	.33	9.80	.01
CMAS	.55	.31	.06	- 1.65	-.32	6.35	.025
Iowa	.60	.36	.05	-14.84	-.20	5.32	.05
NNOM	.63	.39	.03	0.72	.22	4.37	.05