A Microanalysis of Participant Behavior in Familiar and Unfamiliar Test Conditions.

In a microanalysis of the behaviors of examiners and handicapped children during videotaped testing sessions, handicapped students performed better with familiar examiners than with unfamiliar examiners. The children spoke significantly more often and longer when tested by familiar examiners who exercised more frequent and longer intervals of silence, and appeared to use eye contact with examinees as a cue in deciding when to speak. Unfamiliar examiners rarely utilized this cue. Familiar examiners also employed largely directive language in contrast to unfamiliar examiners' speech which was more frequently participatory in nature and longer in duration. The clues to understanding why differences between familiar and unfamiliar examiners' behavior affects test performance, and the implications for testing handicapped children are discussed. Silence is shown to be an important examiner behavior over which test-developers and clinician-trainers should exercise greater control. (Author/CM)
A MICROANALYSIS OF PARTICIPANT BEHAVIOR IN FAMILIAR AND UNFAMILIAR TEST CONDITIONS

Douglas Fuchs, David S. Zern, and Lynn S. Fuchs
The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.
Research Report No. 70

A MICROANALYSIS OF PARTICIPANT BEHAVIOR IN
FAMILIAR AND UNFAMILIAR TEST CONDITIONS

Douglas Fuchs and David S. Zern
Department of Education
Clark University

Lynn S. Fuchs
Institute for Research on Learning Disabilities
University of Minnesota

March, 1982
Abstract

A microanalysis was conducted of the behaviors of examiners and handicapped children during videotaped testing sessions in which handicapped students performed better with familiar than with unfamiliar examiners. The children spoke significantly more often and longer when tested by familiar examiners. Familiar examiners (a) exercised more frequent and longer intervals of silence than unfamiliar examiners, (b) often appeared to use eye contact with examinees as a cue in deciding when to speak whereas unfamiliar examiners rarely utilized this cue, (c) employed largely directive language in contrast to unfamiliar examiners' speech that more frequently was participatory in nature, and (d) spoke for a shorter duration than unfamiliar examiners. These differences between familiar and unfamiliar examiners' behavior provide clues for understanding why differential test performance occurred. Implications for testing handicapped children are discussed.
A Microanalysis of Participant Behavior
in Familiar and Unfamiliar Test Conditions

Over the past 40 years, research has shown that extra-test variables, such as situational and procedural factors, often influence test performance (Masling, 1960; Sattler, 1981; Sattler & Theye, 1967). Similarly, investigation of examiner characteristics has demonstrated that a tester's personal familiarity with select populations, such as young and handicapped children, enhances the examinee's test performance (Fuchs, Featherstone, Garwick, & Fuchs, 1981; Fuchs, Garwick, Featherstone, & Fuchs, 1980; Olswang & Carpenter, 1978; Stoneman & Gibson, 1978). However, previous research appears to regard the influence of examiner familiarity, as well as the effects of other examiner characteristics such as race, sex, personality, and professional experience, as self-explanatory (Masling, 1960; Sattler, 1981); it fails to explore whether, and if so how, familiar and unfamiliar examiners behave differently from one another, why they act as they do, and to what extent, if any, their actions and motives are shaped by examinee's characteristics and test behaviors.

Contributing to the salience of these questions is that many young children currently are tested by unfamiliar adults in school and clinical settings. If the purpose of such testing is to measure and describe children's optimal performance, we need to understand how and why this differential functioning exists before we may hope to vitiate the negative effect that examiner unfamiliarity exerts on certain children's performance.
The present exploratory investigation was a microanalysis of videotaped testing sessions of handicapped children who performed stronger with familiar than unfamiliar examiners within a repeated measures experiment. The purpose of this analysis was to delineate examinee and examiner behaviors that differed between the two experimental conditions in hopes of understanding why differential test performance occurred. Toward this end, the frequency, duration, and quality of examiner and examinee speech and silence were contrasted across familiar and unfamiliar conditions. These facets of the testing situation were selected with the recognition that the examiner-examinee relationship is bi-directional and dynamic. This is in contrast to the view, offered implicitly by much of the prior research on examiner effects, that the examiner-examinee relationship is uni-directional and static (cf. Mehan, 1973, Roth, 1974).

Method

Subjects

Subjects were 15 examinees and six examiners. The examinees were moderately to profoundly speech and/or language handicapped students enrolled in a special education program within a large urban school system in the Midwest. They were a subsample from a previous study (Fuchs, Dailey, & Fuchs, 1982; Fuchs et al., 1980), in which subjects displayed greater descriptive accuracy, verbal fluency, and linguistic complexity in discussing a picture stimulus when tested by a familiar examiner than when tested by an unfamiliar examiner. Examinees were representative of the larger sample in terms of (a) age ($X = 4.7$), (b) sex (60% male), (c) race (20% black and 12% American Indian), (d) length of time that they had been in their present classrooms ($X = 8.1$).
months), and (e) the average direction and magnitude of their different-tail performance in the two examiner conditions.

Familiar examiners (n = 2) were the classroom teachers of the examinees, and, as such, they shared a relatively long-term acquaintanceship with the examinees. Unfamiliar examiners (n = 4) were strangers to the examinees. All examiners were female, certified in early childhood education, and had at least several years of experience working with preschool children in educational settings.

Procedure

Videotaping and editing. The students' test performance with familiar and unfamiliar examiners was videotaped by two AVC 3200 Sony video cameras on one-half inch videotape. The cameras, connected to a Sony 3600 recorder, were placed behind the examiner and examinee. With the aid of a special effects generator (SEG-1), a split screen was created displaying a frontal view of the upper torso and head of both adults and children. With the video cameras, recorder, and special effects generator placed behind blinds, the only visible recording equipment were two table microphones. Examiners were aware of the recording effort; students were not. After the completion of all recording, the videotaped testing sessions were edited electronically with a Panasonic 316C editing tape recorder so that each pupil's recorded performance with familiar and unfamiliar examiners was contiguous.

Experimental task. The videotaped experimental task was similar to certain test items on the Stanford-Binet. It required examinees to describe verbally a high-interest, representational drawing that was selected from Tester's (1966) Teaching Pictures series. An explicit
format governed the administration of this task. Examiners were instructed
to speak only three times: to introduce the task, to prompt the examinee
for additional information, and to communicate termination of the task.
While the content and frequency of examiners' talk was controlled, their
timing, or when during the test session they spoke, was uncontrolled
purposely. (For a more detailed description of the experimental task
and design of the prior study, see Fuchs et al., 1980.)

Videotape analysis. Examiner and examinee behavior was divided
into three general categories: examiner talk (E), examinee talk (C),
and examinee's visual behavior during silent intervals. This last
category was subdivided further into: looking at the stimulus picture (P),
looking at the tester (T), and looking somewhere other than P or T (O).

The videotapes were analyzed by two undergraduate student observers
who did not know the purposes of the study, any of the examiners or
examinees, or whether testing sessions were familiar or unfamiliar.
The observers were required to transcribe the videotapes into the five
behavioral categories, displaying both the temporal sequence of these
behaviors and their duration. The observers were trained to code the
tapes in three one-hour sessions by a graduate student who, like the
observers, knew nothing about the study. Inter-rater reliability was
calculated on 11 of 30 test sessions (15 examinees x 2 conditions), or
37% of the data. Inter-rater agreement ranged from .75 to 1.00, with
an average reliability of .91.

Data Analysis

Examiner and student behaviors were analyzed in terms of both single
behaviors (P, O, T, E) and behavior couplings (e.g., PO, PT, PE).
Chains of three or more behaviors were divided into contiguous couplings by pairing adjacent behaviors in overlapping fashion. A chain of PTPT, for example, was separated into PT, TP, PT. This resulted in 14 behavioral categories, including both single behaviors and couplings. These categories, in turn, provided the frame for a matrix that facilitated the exploration of antecedents to examiner and examinee talk. Since the primary purpose of the study was to identify events that were antecedent to test participants' speech, analyses excluded behavior that followed the last utterance of the child in the test session. A chi square analysis and correlated t tests were used to compare examiner and examinee behavior across familiar and unfamiliar conditions.

Results

Examinee Talk

In the familiar condition, the students spoke more often, \( t(14) = 2.20, p < .05 \), and for greater amounts of time, \( t(14) = 2.37, p < .05 \) (see Table 1). However, when the amount of examinee talk was divided by the frequency of their utterances, no difference was obtained between familiar and unfamiliar conditions (see Table 1).

-------------
Insert Table 1 about here
-------------

Examiner Talk

No statistically significant difference was found between the frequency of familiar and unfamiliar examiner's talk. Yet, as shown
in Table 2, unfamiliar testers spoke for greater amounts of time, \( t(14) = 2.32, p < .05 \). When the amount of examiners' talk was divided by the frequency of their speech, the difference between familiar and unfamiliar testers was not significant.

Insert Table 2 about here

Unfamiliar examiners deviated from test instructions by uttering 11 unsanctioned messages to seven examinees; familiar testers displayed five instances of this rule violation during the testing of three examinees. While the unapproved talk of the familiar examiners seemed to be largely directive (e.g., "Tell me more," "Point to what you mean," "Sit down"), much of the unfamiliar examiners' unsanctioned speech appeared to have a participatory quality (e.g., "Oh yeah?" "Really?" "Hmm!" "Do you want to get up like those girls [in picture]?" "Right. It's [picture] a boy and a doctor.").

Silence

Total silence. Table 3 displays the raw scores for the mean frequency and duration of silent intervals in familiar and unfamiliar conditions. In the familiar situation, silent intervals were more numerous, \( t(14) = 4.99, p < .001 \), and the total duration of silence was greater, \( t(14) = 3.43, p < .01 \).

Insert Table 3 about here
Associated with the higher frequency of silent intervals in the familiar condition was a greater variance ($S^2$: 25.10 vs. 5.21), $t(13) = 4.65$, $p < .01$. With respect to the total duration of silence, there was no significant difference between the variances associated with the two conditions (familiar: $S^2 = 56.68$; unfamiliar: $S^2 = 31.72$). Also, between familiar and unfamiliar situations, there was no significant disparity in the mean duration (duration/frequency) of silent intervals (see Table 3).

Silence before examinee talk. The frequency with which examinee and examiner behavior preceded examinee talk ($C$) was divided by the number of times an examinee spoke, thereby controlling for the greater frequency of $C$ in the familiar condition. Ratios for verbal and non-verbal behavior in the unfamiliar situation then were subtracted from corresponding behavior in the familiar condition. Table 4 displays the findings from these contrasts for individual cases and for the averages across examinees.

As shown in this table, two contrasts, both involving $T$-related behavior (looking at tester), were significant. First, the ratio of $TP$ (looking at tester then at picture) preceding $C$ was greater in the familiar than unfamiliar situation, $t(13) = 2.22$, $p < .05$ (mean difference = + .16). Second, the ratio of $TE$ (looking at tester then examiner talk) that was antecedent to $C$ was greater in the familiar condition, $t(13) = 2.97$, $p = .02$ (mean difference = +.19). Table 4 also indicates that
a combination of PT and TP preceded C more in the familiar situation (mean difference = +.30), whereas, in the unfamiliar condition, PE and EP preceded C more (mean difference = -.22). Additionally, there was a significant relation between examiner condition and T- and E-related behaviors that were antecedent to C, \( \chi^2(1) = 8.98, p < .005. \)

Silence before examiner talk. A total of 73% of familiar examiners' sanctioned speech was preceded immediately by the student looking at the examiner; only 27% of such speech was preceded by the student looking at the stimulus picture. In contrast, for unfamiliar examiners, 40% of sanctioned speech was preceded by the student looking at the examiner and 43% was preceded by the student looking at the picture.

Discussion

A previous study (Fuchs et al., 1981; Fuchs et al., 1980) demonstrated that handicapped preschool children employed (a) more words, (b) more complex syntactic and semantic structures, and (c) greater descriptive accuracy in discussing a picture when tested by familiar examiners than when tested by unfamiliar examiners. Employing a representative sub-sample of the subjects involved in that study, this investigation re-analyzed the children's verbal fluency by examining the duration as well as the frequency of their speech, and further by locating their talk in relation to their visual behavior during silent intervals and to examiner speech. This microanalysis was undertaken to identify the antecedents to examinee speech in familiar and unfamiliar conditions in order to
Examiner talk, a logical antecedent to children's speech in a test situation, appeared to be unrelated to the examinees' greater fluency in the familiar condition. Familiar examiners did not speak more frequently than unfamiliar examiners; in fact, the total duration of familiar examiners' talk was less than that of unfamiliar examiners' speech. However, both the frequency of silent intervals and the total duration of silence preceding examinee talk were associated positively with examinees' more frequent and extended verbal production in the familiar situation.

The present study suggested that examiner behavior contributed to the greater frequency of silent intervals observed in the familiar situation. Analysis of apparent cues used by examiners in the determination of when to speak revealed that familiar examiners were significantly more likely than unfamiliar examiners to speak only after examinees had looked up from the stimulus picture to look at them; unfamiliar examiners displayed a tendency to speak more often than familiar examiners while the children still were viewing the stimulus. These disparate patterns of examiner behavior were reflected by the fact that TP and PT couplings preceded C 2.3 times more frequently in the familiar condition, whereas EP and PE couplings were antecedent to C 1.7 times more often in the unfamiliar condition. Similarly, the pattern TPC--examinee looking at the tester, at the stimulus picture, and then speaking--was significantly more frequent in the familiar condition.
Two points need to be made both with respect to the findings and interpretations developed from them. First, due to the relatively large number of correlated t tests used and the inter-relatedness of much of the data, findings may reflect Type I error. Second, this was a correlative study rather than an experimental investigation; demonstrated relationships between examiner and examinee behavior are not necessarily causative.

Nevertheless, findings were consonant with the supposition that the greater amount of and more frequent silence preceding examiner talk in the familiar condition contributed to examinees' greater verbal fluency in the familiar situation. Findings also suggested that examiner behavior was at least partly responsible for this silence. The data from this study are insufficient to explain the obtained differences in the behavior of familiar and unfamiliar examiners. It is reasonable to speculate, however, that the unfamiliar examiners' relatively infrequent use of silence (i.e., their apparent proneness to terminate the testing in an untimely manner) was related to their ignorance of examinees' skill level and knowledge base. Confronted by examinees' silence, unfamiliar examiners had to decide whether to ignore this silence and encourage continued effort or to withdraw the test demand. Presumably, this decision normally is based on an understanding of the examinees' capabilities. Because unfamiliar examiners in this study had minimal information about examinees' ability levels, were presumably perceptive about and empathetic towards children's feelings, and were no doubt aware of their need for examinees' cooperativeness, they may have believed
there was no alternative but to behave conservatively and employ examinees' silence as a primary cue in determining when to conclude testing.

Consonant with the speculation that unfamiliar testers were comparatively uneasy during assessment were findings that they spoke to examinees for a significantly greater duration of time than familiar testers and that their relatively more frequent unsanctioned speech appeared to be participatory in nature.

In contrast to unfamiliar testers, familiar examiners, by definition, had a more accurate notion of examinees' capabilities. When testing children who were reticent, yet known to be capable, familiar testers appeared comparatively unresponsive to examinee silence, and supposively communicated an expectation that the children demonstrate their potential. In agreement with this interpretation, familiar examiners were more variable in their deployment of silence than were unfamiliar testers. Also, because familiar testers shared a relatively long acquaintanceship with the examinees, the familiar examiners may have been less concerned than unfamiliar examiners about examinees' discontinued cooperation during assessment.

If testers' differential use of silence at least partially explains why select groups of children tend to perform more strongly with familiar than with unfamiliar examiners, then silence is an important examiner behavior over which test developers and clinician-trainers should exercise greater control. Future research might use experimental designs to explore the importance of silence to test performance. On the basis of this investigation, one might hypothesize that if familiar and unfamiliar examiners were required to employ silence similarly, the negative effects of examiner unfamiliarity may be vitiated.
At a more general level, familiar and unfamiliar examiners' differential use of silence (as well as other differences in participants' behavior between familiar and unfamiliar conditions) corroborates the view that examiners and examinees interact in a bi-directional, dynamic, and creative process; in a sense, a child's test performance is a social accomplishment to which both examinee and examiner may lay claim (see Mehan, 1978). Simultaneously, findings from this study question the still popular Galtonian view that a test is no more or less than a sample of the subject's responses to a standardized non-personal stimulus. The present study and past investigations suggest that such a view may interfere with both securing optimal test performance and developing meaningful educational programs that are based upon formal assessment results.
References


MacEachern, D. Personal communication, February 18, 1982.


Douglas Fuchs also is a Postdoctoral Fellow at the Institute for Research on Learning Disabilities.

The following formula (MacEachern, 1982) was used for testing the difference between two variances that were calculated on the same sample:

\[
 t = \frac{(s_1^2 - s_2^2) \sqrt{N-2}}{2s_1s_2 \sqrt{1-r^2}}
\]
Table 1

Amounts of Examinee Talk with Familiar and Unfamiliar Examiners<sup>a</sup>

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Examiner Condition</th>
<th>Familiar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency*</td>
<td></td>
<td>3.80(2.43)</td>
<td>2.33(1.29)</td>
</tr>
<tr>
<td>Duration*</td>
<td></td>
<td>9.53(8.68)</td>
<td>5.83(4.90)</td>
</tr>
<tr>
<td>Duration/Frequency</td>
<td></td>
<td>2.60(2.40)</td>
<td>2.35(1.68)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Entries are means and standard deviations (in parentheses) of the number of times the 15 examinees spoke (Frequency) and the number of seconds they spoke (Duration).

*Difference between Familiar and Unfamiliar conditions was statistically significant (p < .05).
Table 2

Amounts of Examiner Talk in Familiar and Unfamiliar Conditions

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Examiner Condition</th>
<th>Familiar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>1.33(.82)</td>
<td>1.80(1.15)</td>
<td></td>
</tr>
<tr>
<td>Duration*</td>
<td>1.28(.92)</td>
<td>2.71(1.91)</td>
<td></td>
</tr>
<tr>
<td>Duration/Frequency</td>
<td>.90(.25)</td>
<td>1.56(1.01)</td>
<td></td>
</tr>
</tbody>
</table>

*a Entries are means and standard deviations (in parentheses) of the number of times examiners spoke (Frequency) and the number of seconds they spoke (Duration).

*b Familiar examiners N = 2; Unfamiliar examiners N = 4.

* Difference between Familiar and Unfamiliar conditions was statistically significant (p < .05).
Table 3
Silence in Familiar and Unfamiliar Conditions\textsuperscript{a}

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Examiner Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Familiar</td>
<td>Unfamiliar</td>
<td></td>
</tr>
<tr>
<td>Frequency***</td>
<td>9.60(4.69)</td>
<td>5.07(2.28)</td>
<td></td>
</tr>
<tr>
<td>Duration**</td>
<td>13.78(7.53)</td>
<td>6.01(5.63)</td>
<td></td>
</tr>
<tr>
<td>Duration/Frequency</td>
<td>2.17(.57)</td>
<td>2.24(1.29)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Entries are the means and standard deviations (in parentheses) of the number of instances of silence (Frequency) and their length in seconds (Duration).

**Difference between Familiar and Unfamiliar conditions was statistically significant (p < .01).

***Difference between Familiar and Unfamiliar conditions was statistically significant (p < .001).
Table 4

Ratios of Behavior Preceding Examinee Talk for Familiar Minus Unfamiliar Conditions\(^a\)

<table>
<thead>
<tr>
<th>Examinees(^b)</th>
<th>Behavior Categories(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P O T E PO OP PT TP PE EP OE EO TE ET</td>
<td></td>
</tr>
<tr>
<td>1 +.25 - .50 -.50 +.75 +.50 -.50 +.25</td>
<td>+.25 -.50</td>
</tr>
<tr>
<td>2 +.75</td>
<td>-1.00 -.75</td>
</tr>
<tr>
<td>3 - .25</td>
<td></td>
</tr>
<tr>
<td>4 +.25 +.25 +.25</td>
<td>+.25 -.25</td>
</tr>
<tr>
<td>5 -.24 +.38 +.29</td>
<td>-.19 +.25</td>
</tr>
<tr>
<td>6 +.25 - .50</td>
<td>+.50</td>
</tr>
<tr>
<td>7 -.50</td>
<td>+.50 +.50</td>
</tr>
<tr>
<td>8 +.25 - .25</td>
<td>-.25 -.25 -.50 -.25 +.50</td>
</tr>
<tr>
<td>9 -.21</td>
<td>-.36 +.57 +.14 -.36</td>
</tr>
<tr>
<td>10 +.67</td>
<td>-.50 +.33</td>
</tr>
<tr>
<td>11 -.50</td>
<td>+1.00 +.50 +.50 +.50</td>
</tr>
<tr>
<td>12 .00</td>
<td>.00</td>
</tr>
<tr>
<td>13 +.20 +.10 +.10</td>
<td>-.33 -.33 -.57 +.10 +.10 -.23</td>
</tr>
<tr>
<td>14 -.25</td>
<td>+.50 +.50 -.25 +.25 +.25</td>
</tr>
</tbody>
</table>

| \(\bar{x}\) | +.05 +.03 -.02 -.03 -.04 -.04 +.14 +.16* -.19 -.03 -.03 -.01 +.19** -.02 |
| SD | .39 .10 .07 .14 .13 .13 .38 .27 .52 .43 .14 .07 .24 .18 |

\(^a\) Frequency of each behavior was divided by the frequency of examinee talk. Spaces in table reflect absence of behavior preceding examinee talk and are treated as zeroes in calculating means and standard deviations.

\(^b\) One examinee was excluded because of no talk in the unfamiliar condition.

\(^c\) P = Examinee looks at picture; O = Examinee looks at something other than picture or tester; T = Examinee looks at tester; E = Examiner talk.

*Difference between Familiar and Unfamiliar conditions was statistically significant (\(p < .05\)).

**Difference between Familiar and Unfamiliar conditions was statistically significant (\(p < .01\)).
PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.


Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.


Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.


Epps, S., McCue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McCue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.


