Journal of Applied School Psychology, 29:37–51, 2013

ISSN: 1537-7903 print / 1537-7911 online
DOI: 10.1080/15377903.2013.751475

Routledge
Taylor & Francis Group

Examiner Familiarity Effects for Children With Autism Spectrum Disorders

JULIA E. SZARKO
The Pennsylvania State University, State College, Pennsylvania, USA

ALEC J. BROWN
Arizona State University, Arizona, USA

MARLEY W. WATKINS
Baylor University, Waco, Texas, USA

*The authors examined the difference in standardized test performance when familiar versus unfamiliar examiners tested 26 preschool and elementary-aged children with autism. The children were matched by age, severity, and developmental level and then randomly placed into familiar and unfamiliar examiner groups. Familiarity with the examiner was established before test administration for children in the treatment group. Both groups were administered 2 subscales of the Psychoeducational Profile-Revised. There was a statistically significant difference in favor of the children tested by the familiar examiners on the cognitive verbal subscale (*d = *.43) and on the cognitive performance subscale (*d = *.47), indicating that examiner familiarity had positive effects on the test performance of children with autism. Given these results, it appears that professionals who are responsible for administering standardized tests to children with autism should make pretest contact with these children to reduce test procedure bias.*

Keywords autism, familiarity, test administration, test bias, assessment, validity

---

Julia E. Szarko is now at the Central Bucks School District, Doylestown, Pennsylvania.

This report is based on the dissertation of the first author that was supervised by Professors John Salvia, Joseph French, Robert Hale, and Robert Stevens. The contributions of professors Salvia, French, Hale, and Stevens are gratefully acknowledged.

Address correspondence to Marley W. Watkins, Baylor University, 1 Bear Place #97301, Waco, TX 76798, USA. E-mail: marley_watkins@baylor.edu

It is critical to ensure the validity of test-related inferences (American Educational Research Association, American Psychological Association, & National Council in Measurement in Education, 1999). This goal may be particularly challenging when testing children who have autism (Ingersoll, 2011). In the past, some researchers have suggested that these children were untestable (Clark & Rutter, 1979). Although researchers and practitioners now recognize the inaccuracy of this broad generalization, it is apparent that there are unique challenges in the assessment of this population (El-Ghoroury & Krackow, 2012; Ozonoff, Goodlin-Jones, & Solomon, 2005; Volkmar, Chawarska, & Klin, 2005). Disruptive behaviors (e.g., self-stimulation, task avoidance), impaired attention, impaired social interaction, and communication deficits can all influence a child's performance on standardized tests, meaning that "standardized testing may be measuring the child's test-taking disability rather than intellectual or verbal ability" (L. K. Koegel, Koegel, & Smith, 1997, p. 241).

Examiners can attempt to minimize the influence of interfering behaviors by structuring and adapting the testing situation to meet the child with autism's atypical response style (Egel, Holman, & Barthold, 2011). There are a variety of considerations when conducting assessments of students with autism (Filipek et al., 2000; Handleman, 1992; Marcus, Lansing, & Schopler, 1993; Matson, Beighley, & Turygin, 2012; Ozonoff et al., 2005; Volkmar et al., 2005). In general, examiners could use multiple sources, including interviews with teachers and parents, to gather information about response modes, communication levels, preferred reinforcers, and idiosyncratic strengths and weaknesses. This information can then be used to adapt the assessment to the needs of the child. For example, an examiner may modify the presentation of the material (shorten directions, use dramatic gestures), the response mode (allow the child to point), or methods of reinforcement for on-task effort. However, it is important to recognize that modifications that are not part of the standard administration will invalidate norm-referenced interpretations.

Factors such as motivation, attention, and communication are not the only barriers to valid assessment. Contextual factors such as test location, examiner personality, examiner error, and examinee understanding of the purpose of the test can also influence a child's test performance, making it important to consider the effect of these variables (Fuchs & Fuchs, 1986; Kuentzel, Hetterscheidt, & Barnett, 2011; Waterman, McDermott, Fantuzzo, & Gadsden, 2012). One contextual factor that has been studied in a variety of child populations is examiner familiarity. Some populations show differential performance on standardized assessments when tests are administered by familiar, as opposed to unfamiliar, examiners (Fuchs, Fuchs, Power, & Dailey, 1985; Kinnie & Sternlof, 1971; Olswang & Carpenter, 1978; Stoneman & Gibson, 1978). For example, Black and Hispanic children scored significantly higher on standardized tests administered by familiar, as opposed to unfamiliar, examiners, whereas Caucasian students performed similarly under both conditions (Fuchs & Fuchs, 1989).

Similar studies conducted with children with disabilities have more direct implications for autism. The majority of these studies have been conducted with children with language disorders (Fuchs, Featherstone, Garwick, & Fuchs, 1984; Fuchs, Fuchs, Dailey, & Power, 1985; Fuchs, Fuchs, Garwick, & Featherstone, 1983). The results of these studies have consistently indicated that children with language disorders scored significantly lower on standardized tests when they were administered by an unfamiliar examiner, whereas the performance of their nondisabled peers was unaffected by examiner familiarity. However, examiner familiarity cannot be assumed to affect all populations in the same way. Fuchs, Fuchs, and Power (1987) studied the effect of examiner familiarity on language performance in two populations, children with learning disabilities and children with intellectual disabilities. The results indicated that children with learning disabilities performed significantly better when the standardized test was administered by a familiar examiner, as opposed to an unfamiliar examiner. On the other hand, children with intellectual disabilities performed similarly in both conditions.

Fuchs and Fuchs (1986) argued that if situational factors in the testing environment systematically increase or decrease the performance of certain groups of examinees, such factors could be considered systematic sources of error or *test procedure bias*. This type of systematic error could have serious implications for the validity of diagnostic and programmatic decisions based on standardized test results. Considering the workload and time constraints experienced by many school psychologists, it may not be practical to spend significant time interacting with each examinee before assessment. However, if there are identified subgroups that are known to "perform suboptimally with an unfamiliar examiner, then examiners might establish pretest contact with that subgroup of pupils" (Fuchs et al., 1987, p. 196).

Children with autism represent one subgroup for whom the examiner familiarity effect has not been adequately examined. Given that children with autism experience interpersonal weaknesses, difficulties with novel situations, deficits in joint attention, and communication problems (Johnson, Myers, & the Council on Children with Disabilities, 2007), it is reasonable to expect that they may perform better when tested by familiar, as opposed to unfamiliar, examiners. Therefore, the purpose of this study was to explore the possibility of test procedure bias due to examiner familiarity effects for students with autism.

METHOD

Participants

Participants were 26 children (19 boys, 7 girls) who ranged in age from 48 months to 88 months (M = 64.5 months, SD = 10.8 months). Ethnic background of the participants was 65% White, 23% Black, and 12% Asian-Pacific. The participants were enrolled in four preschool special education

classrooms and two Kindergarten special education classrooms in a southeastern metropolitan school district. All participants were diagnosed as having autism or atypical pervasive developmental disorder as their primary disability by school teams. Children with intellectual disabilities as documented in school records were excluded from the study.

Measures

Psychoeducational Profile-Revised Subtests

All participants were administered the Cognitive Performance and Cognitive Verbal subtests of the Psychoeducational Profile-Revised (Schopler, Reichler, Bashford, Lansing, & Marcus, 1990), which is an inventory of behaviors and skills designed to identify uneven and idiosyncratic learning patterns for children with autism who were 6 months to 7 years old. The profile depicts a child's relative strengths and weaknesses in different areas of development. The Psychoeducational Profile-Revised cognitive subscales have demonstrated internal consistency reliabilities of .94 to .97 (Althouse, 1996; Villa et al., 2010), interrater reliabilities of .88 to .98 (Bock & Hurlbutt, 2001—2002; Villa et al., 2010), and are strongly correlated with the composite IQ score on the Stanford-Binet Intelligence Scales, Fourth Edition ($r = .73$; Delmolino, 2006) and domain scores on the Italian version of the Vineland Adaptive Behavior Scale ($r = .87$; Villa et al., 2010).

Behavioral Observations

In addition to standardized test scores, observations were used during test sessions to gather information about three behavioral variables. All test sessions were videotaped and then reviewed by six independent observers who coded three target behaviors: (a) the number of prompts provided by the examiner, (b) the number of items refused by the child, and (c) the number of atypical behaviors (including, but not limited to: echolalia, biting, smelling objects, waving objects in front of their eyes, and screaming) exhibited by the child. The coders were all state certified school psychologists employed in the school district. The coders were unaware of the experimental condition to which the children had been assigned (familiar/unfamiliar). The first author trained the coders to mastery during a 2-hr session where they practiced coding the video of a participant not used in the study. Two independent raters coded 20% of the test session videotapes with interrater reliability of .90.

Procedure

Six examiners (5 female, 1 male; all 6 White) were recruited to administer the Psychoeducational Profile-Revised to the participants. Four of the examiners

were undergraduate students studying psychology and two were graduate students in psychology. The examiners were individually trained to mastery in administration of the Psychoeducational Profile-Revised subtests by the first author. The examiners were also instructed on how to build rapport with examinees. Suggestions on building rapport included using short verbal phrases, engaging in play with objects preferred by the child, and avoiding physical touch. The examiners were not told the purpose of the experiment nor the research hypotheses.

Following institutional review board approval, the parents of all children with autism or atypical pervasive developmental disorder diagnoses in six preschool and Kindergarten special education classrooms were asked to allow their child to participate in this study. Twenty-six parents consented. No pretests were administered, but the six teachers rated their students on language functioning (verbal or nonverbal), severity of autism (severe, moderate, or mild), and instructional level (above grade level, at grade level, or below grade level). After being matched on these three variables (100% of pairs matched on language functioning and instructional level, 92% of pairs matched on severity of autism), children were also matched by age (median deviation in age of paired children was five months), ethnic background (62% matched), and gender (62% matched). A detailed description of matched pairs is provided in Table 1. Matched pairs were then randomly assigned to the familiar ($n = 13$) or unfamiliar ($n = 13$) group in a posttest-only control group design. In the unfamiliar condition examiners had no interpersonal interaction with the examinee before the test session. In the familiar condition the examiner interacted with the examinee in non–test situations until indicators of personal familiarity were exhibited by the child.

Indicators of personal familiarity were individually identified for each child in the familiar examiner condition group. Each child's teacher was interviewed and asked to describe behaviors that the child typically demonstrated only after becoming familiar with an adult. Table 2 summarizes the behaviors reported by the teachers. Although the behaviors indicative of familiarity varied considerably among the children, the familiar behaviors most commonly reported by their teachers were initiating verbal contact, smiling, and initiating physical contact. On the other hand, some of the familiarity indicators were more idiosyncratic and unique, such as asking for his mother or failing to cry. The number of familiar behaviors designated for each child ranged from 2 to 4 ($M = 3.5$ behaviors).

To establish familiarity with an examinee in the familiar examiner condition, the examiner spent time in the classroom during work, recreation, and snack activities. During these activities the examiner interacted specifically with the target child. After the interaction period, the examiner rated whether the child had demonstrated the preselected familiar behaviors. Only when the child had exhibited all of the preidentified behaviors was familiarity considered to be established. The amount of time spent establishing familiarity

TABLE 1 Participant Matches on Age, Gender, Ethnicity, Language Functioning, Severity of Autism, and Instructional Level

	Familiar examiners						Unfamiliar examiners					
Pair	Age	Gender	Ethnicity	Language functioning	Severity autism	Instruction level	Age	Gender	Ethnicity	Language functioning	Severity autism	Instruction level
14–8	5–4	Male	White	Verbal	Mild	At grade level	5–6	M	White	Verbal	Mild	At grade level
11–4	4–1	Male	White	Nonverbal	Severe	Below grade level	7–4	M	White	Nonverbal	Severe	Below grade level
9–22	5–5	Male	Asian	Verbal	Mild	At grade level	5–8	M	Black	Verbal	Mild	At grade level
24–23	6–7	Male	Black	Verbal	Moderate	Below grade level	6–6	M	White	Verbal	Moderate	Below grade level
13–27	4–5	Male	White	Verbal	Mild	At grade level	4–6	M	White	Verbal	Mild	At grade level
19–20	4–0	Female	White	Verbal	Severe	Below grade level	4–4	M	Asian	Verbal	Severe	Below grade level
15–10	5–8	Male	White	Verbal	Severe	Below grade level	4–7	M	White	Verbal	Severe	Below grade level
7–2	6–3	Female	White	Verbal	Mild	At grade level	5–10	F	White	Verbal	Mild	At grade level
5–18	6–1	Female	Black	Verbal	Moderate	Below grade level	5–5	M	Black	Verbal	Moderate	Below grade level
25–16	5–1	Female	Black	Nonverbal	Moderate	Below grade level	4–0	M	White	Nonverbal	Severe	Below grade level
26–6	5–4	Female	Black	Verbal	Mild	At grade level	5–7	M	Asian	Verbal	Mild	At grade level
21–12	4–4	Male	White	Verbal	Severe	Below grade level	5–3	M	White	Verbal	Severe	Below grade level
1–3	6–8	Female	White	Verbal	Moderate	Below grade level	6–0	M	White	Verbal	Moderate	Below grade level

Note. Pair prefers to familiar participant-unfamiliar participant. Age is expressed in years-months.

TABLE 2 Time and Number of Visits Needed for 13 Familiar Group Participants to Exhibit Behaviors That Established Familiarity

Participant	Familiar behaviors	Time (hours)	Number of visits
3	• initiates verbal contact • smiles • talks to you	4	2
5	• smiles • initiates physical contact • says hello • keeps hands off face when sees you	6	3
7	• initiates verbal contact • smiles • initiates physical contact • no TV talk while working with you	6	3
9	• initiates verbal contact • smiles • initiates physical contact • makes eye contact	4	2
11	• initiates physical contact (e.g., sits in your lap) • body relaxes	8	4
13	• smiles • talks to you • engages you in activities • does not cry or ask for mother	6	3
14	• initiates verbal contact • talks to you • allows you to interact in activity	4	3
15	• no TV talk while working with you • physical exchange during play • gives you space to join in during play • does not scream	11	5
19	• initiates verbal contact • initiates physical contact • initiates play	4	4
21	• initiates physical contact • sits beside you • does not scream	4	2
24	• allows you to interact in activity • shows affection (e.g., hugs you) • calls you by name • asks for help	2	1
25	• initiates verbal contact • smiles • initiates play	1	1
26	• initiates verbal contact • smiles •talks to you • sits beside you	2	1

with the children before testing ranged from approximately 1 to 11 hr (M = 4.8). The number of visits ranged from 1 to 5 (M = 2.6). Table 2 provides results for each familiar examiner group participant.

After familiarity had been established in non–test environments, test sessions were conducted for both groups in a small office or screened area of an empty classroom. All examiners, regardless of experimental condition, engaged in a brief warm-up or rapport building period with the children before testing. However, no attempt was made to familiarize the student with the setting or materials beyond standard rapport building procedures.

The examination of familiar and unfamiliar children was counterbalanced for examiners. For example, Examiner A tested an unfamiliar participant first and then a familiar participant, and Examiner B tested a familiar participant and then an unfamiliar one, and so on. The examiners administered the test items from the Psychoeducational Profile-Revised in the order in which they appeared in the manual, following standardized procedures. Sessions lasted 16 to 65 minutes (M = 45.2) and each participant was videotaped while the examiner administered the items from the test. Following the test session, the scores from the Psychoeducational Profile-Revised subtests were calculated. The six independent coders then independently reviewed the videotapes and rated the children's test session behavior.

RESULTS

Psychoeducational Profile-Revised

Scores on the Psychoeducational Profile-Revised were the dependent variable of primary interest in this study. Table 3 displays the mean cognitive verbal and cognitive performance developmental age scores (represented in months) for the familiar and unfamiliar groups. The mean scores for the

TABLE 3 Test Behaviors and Performance of Participants on the Psychoeducational Profile-Revised Cognitive Verbal and Cognitive Performance Subtests Administered by Familiar (n = 13) and Unfamiliar (n = 13) Examiners

	Familiar examiners		Unfamiliar examiners		
	M	*SD*	*M*	*SD*	*d*
Test behaviors					
Refusals	3.4	5.0	4.8	6.3	.25
Prompts	1.5	1.5	1.2	0.8	.25
Stereotypes*	0.2	0.2	0.4	0.4	.63
Cognitive scale					
Verbal**	13.5	6.9	10.6	6.5	.43
Performance*	16.5	6.5	13.3	7.0	.47

*$p < .05$. **$p < .01$.

familiar group were higher than the mean scores for the unfamiliar group on the cognitive verbal and the cognitive performance sections of the Psychoeducational Profile-Revised. Standardized mean effect sizes were .43 and .47, which approached Cohen's (1988) designation of medium size, and were almost identical to previously studies with relatively long duration ($d = .48$) and demanding ($d = .43$) tests (Fuchs & Fuchs, 1986).

To ascertain whether the differences observed in the sample could be inferred to the population, a Wilcoxon Matched-Pairs Signed-Ranks Test was performed. The Wilcoxon test was selected for this analysis because it (a) is appropriate for studies with matched-subjects designs, (b) does not assume a normal distribution, and (c) is relatively powerful for small samples (Blair & Higgins, 1985; Fahoome, 2002). Given that previous research was consistent in finding positive or neutral effects, a one-sided test at the .05 level was selected (Nolan & Heinzen, 2011). The resulting test indicated significant differences between the familiar and unfamiliar groups on the cognitive verbal subscale, $z = -2.51$, $p = .006$, and the cognitive performance subscale, $z = -1.79$, $p = .037$. These results indicate that the group tested by familiar examiners scored higher on both subscales than the group tested by unfamiliar examiners.

Behavioral observation results

Independent coders observed video recordings of the test sessions and recorded data on the number of refusals, prompts, and atypical behaviors. Means and standard deviations were computed for each group. These results are also reported in Table 3. To determine if the differences observed in the sample could be inferred to the population, three separate Wilcoxon Matched-Pairs Signed-Ranks Tests were performed. The results did not indicate a significant difference between the familiar examiner and unfamiliar examiner groups in the number of prompts given, $z = -.94$, $p = .17$, or the number of items refused by the child, $z = -.99$, $p = .16$. However, there was a significant difference between the two groups for the atypical behavior variable, $z = -1.65$, $p = .05$. Although this difference was statistically significant, the mean number of behaviors was so low ($M = 0.2$ for the familiar group, $M = 0.4$ for the unfamiliar group) for both groups that its practical significance is doubtful.

DISCUSSION

The results of the present study indicated that there was an examiner familiarity effect for children with autism. Children who were tested by a familiar examiner scored higher on the Psychoeducational Profile-Revised cognitive

subtests than children who were tested by an unfamiliar examiner ($d = .43$ to .47). In terms of a normative test with mean of 100 and standard deviation of 15, a familiar examiner would raise the typical score from 100 to around 107. The standard error of measurement for this same scale would be approximately three points so an increase of seven points would be practically significant. These results are similar to those that have been found for children with language disorders and children with learning disabilities (Fuchs & Fuchs, 1986; Fuchs et al., 1983, 1984, 1987; Fuchs, Fuchs, Dailey, & Power, 1985; Fuchs, Fuchs, Power, & Dailey, 1985), indicating that the implications of previous research may be extended to children with autism.

This study did not include a sample of peers without disabilities; however, previous research has indicated that examiner familiarity may not affect standardized test performance in the same way for children without disabilities as it does for children with disabilities. As such, examiner familiarity appears to constitute a source of systematic bias called test procedure bias (Fuchs & Fuchs, 1986) or construct-irrelevant variance (Messick, 1995). The present study provides preliminary evidence that test procedure bias may pose a threat to the validity of standardized test inferences for the population of young children with autism.

There are several plausible explanations for the familiarity effect. One hypothesis is that certain features of autism, such as a strong need for routine and difficulty adapting to change, may account for the familiarity effect because working with the familiar examiner may be less of a deviation from the child's routine. Other possible explanations include differences in motivation, fewer disruptive behaviors, and decreased anxiety.

Some authors have argued that familiar examiners may be more motivating, especially for children with autism (L. K. Koegel et al., 1997; R. L. Koegel & Mentis, 1985). In contrast, motivation to respond may be reduced and performance may be negatively affected when a test is administered by an unfamiliar examiner. If this hypothesis is correct, one might expect that children tested by unfamiliar examiners would refuse to complete more items and would require more prompts than those tested by a familiar examiner. However, in the present study the familiar examiner and unfamiliar examiner groups did not differ significantly on the number of refusals or prompts, indicating that motivational differences may not be the best explanation for the familiarity effect.

Another explanation might be that a reduction in interfering behaviors accounts for the differences between groups. It has been proposed that disruptive behaviors can present a significant barrier to valid test administration with children who have autism (L. K. Koegel et al., 1997). Furthermore, there is some evidence that children with autism exhibit increased stereotypic behaviors when task demands are high (Clark & Rutter, 1979; Durand & Carr, 1987), and when they are interacting with unfamiliar individuals (Charlop, 1986; Runco, Charlop, & Schreibman, 1986). Consistent with this hypothesis,

the group of children tested by unfamiliar examiners engaged in significantly more atypical behaviors than children tested by familiar examiners. This finding provides some support for the idea that a decrease in stereotypical behavior may account for the familiar examiner group's higher scores on the Psychoeducational Profile-Revised. However, this interpretation should be made cautiously because the mean number of atypical behaviors was low in both groups (0.2 and 0.4, respectively), and it seems unlikely that such a small difference sufficiently accounts for differences in test performance.

Anxiety provides another potential explanation for the results of the present study. Anxiety is mentioned as one of the biobehavioral patterns associated with autism by Mesibov and Shea (n.d.). Groden, Cautela, Prince, and Berryman (1994) have theorized that interpersonal interactions can be especially anxiety producing for children with autism. In theory, the more familiar and predictable a situation is for individuals with autism, the more comfortable and less anxious they will feel. In typically developing individuals, anxiety has been shown to affect processing efficiency and, to a lesser extent, response accuracy by influencing working memory and central executive functions related to attention (Eysenck, Derakshan, Santos, & Calvo, 2007). Given the influence of anxiety on test performance in general, and the possibility that the unfamiliar examiner condition may be anxiety producing for children with autism, anxiety could account for the results of the present study.

It is also possible that a combination of factors accounted for differential test performance. Some researchers have suggested that stereotypic behaviors are one expression of anxiety in individuals with autism (Cairns, 1986; Durand & Carr, 1987; Groden et al., 1994). Therefore, the small difference in atypical behaviors observed in the present study could be indicative of greater anxiety in the unfamiliar examiner group, which could have influenced cognitive functioning during testing. Additional research will be needed to clarify the factors responsible for the examiner familiarity effect.

As with all field-based research, there were several limitations to the present study. First, the study sample was small and diverse; the children ranged from low to high functioning and from mild to severe in autistic characteristics. Given the small sample size, the posttest-only control group design may have failed to control for preexisting group differences. Matching the children before random group assignment provided greater equivalence between the two experimental conditions, but did not allow for the evaluation of the effect of familiarity by severity of characteristic and degree of functioning. Researchers with larger or more focused samples of children with autism might modify or strengthen the results found in the present study. For example, there might be an interaction between level of functioning and test performance.

The characteristics of the Psychoeducational Profile-Revised are also important. The Psychoeducational Profile-Revised is a developmental test

designed with the characteristics of autism in mind (Schopler et. al., 1990). The tasks are designed to be visually interesting and stimulating to children with autism and the format of the test is more flexible than traditional standardized tests. These Psychoeducational Profile-Revised characteristics should produce fewer examiner effects than many tests, but caution should be exercised in generalizing the findings of the present study to other tests, including the Psychoeducational Profile: Third Edition, until further evidence is obtained. Researchers might investigate the extent to which there is differential testing performance across various standardized tests.

Last, the examiners in the current study were novices and had little or no experience testing children with autism before the experiment. Research conducted by Fuchs, Fuchs, Dailey, and Power (1985) found that children with speech and/or language disorders performed significantly better with familiar than unfamiliar examiners, regardless of examiner experience. Nevertheless, it is possible that more experienced examiners who were familiar with autism could have better overcome the effects of unfamiliarity.

The results of the present study provide preliminary evidence for extending the results of previous examiner familiarity research to the population of young children with autism. Regardless of the explanation for these findings, there are implications for practice. If examiner familiarity is responsible for systematic bias in the assessment of children with autism, then precautions should be taken to guard against this bias. On the basis of the results of the present study, it would seem that individuals who are responsible for administering standardized tests to children with autism should make pretest contact with these children. Although the optimal length and nature of interaction is at present unknown, research with other populations has indicated that a minimum of 1 to 2 hr of contact is necessary (Fuchs et al., 1987). Finding additional time to interact with children before assessment may be difficult; however, creative solutions may exist. For example, if an examiner were responsible for evaluating several children from the same classroom throughout the year, the examiner could spend time interacting with multiple children during a group activity. Another option might be to have the interaction time serve a dual purpose of pretest contact and observation of the child's behavior. In this way the overall quality of the evaluation may be doubly increased by obtaining unbiased test results and by simultaneously gaining a richer understanding of the child through extended direct observation in the natural environment.

REFERENCES

Althouse, J. (1996). *An investigation of the reliability and validity of the Psychoeducational Profile - Revised*. Unpublished master's thesis, The Pennsylvania State University, State College.

American Educational Research Association, American Psychological Association, National Council in Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the paired samples *t* test to that of Wilcoxon's signed-ranks test under various population shapes. *Psychological Bulletin*, *97*, 119–128.

Bock, M. A., & Hurlbutt, K. (2001–2002). Preacademic and vocational assessment: The key to effective educational programming for students with autism. *Assessment for Effective Intervention*, *27*, 81–88.

Cairns, R. (1986). Social development: Recent theoretical trends and relevance for autism. In E. Schopler & G. Mesibov (Eds.), *Social behavior in autism* (pp. 15–30). New York, NY: Plenum Press.

Charlop, M. H. (1986). Setting effects on the occurrence of autistic children's immediate echolalia. *Journal of Autism and Developmental Disorders*, *16*, 473–483.

Clark, E. V., & Rutter, M. (1979). Task difficulty and task performance in autistic children. *Journal of Child Psychology and Psychiatry*, *20*, 271–285.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Delmolino, L. M. (2006). Use of DQ for estimating cognitive ability in young children with autism. *Journal of Autism and Developmental Disorders*, *36*, 959–963.

Durand, V. M., & Carr, E. G. (1987). Social influences on "self-stimulatory" behavior: Analysis and treatment application. *Journal of Applied Behavior Analysis*, *20*, 119–132.

Egel, A., Holman, K., & Barthold, C. (2011). *School success for kids with autism*. Waco, TX: Prufrock Press.

El-Ghoroury, N. H., & Krackow, E. (2012). Enhancing the identification of autism spectrum disorders via a model of culturally sensitive childhood assessment. *Professional Psychology: Research and Practice*, *43*, 249–255.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*, 336–353.

Fahoome, G. (2002). Twenty nonparametric statistics and their large sample approximations. *Journal of Modern Applied Statistical Methods*, *1*, 248–268.

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Dawson, G., . . . Volkmar, F. R. (2000). Practice parameter: Screening and diagnosis of autism. *Neurology*, *55*, 468–479.

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

Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research*, *56*, 243–262.

Fuchs, D., & Fuchs, L. S. (1989). Effects of examiner familiarity on Black, Caucasian, and Hispanic children: A meta-analysis. *Exceptional Children*, *55*, 303–308.

Fuchs, D., Fuchs, L. S., Dailey, A. M., & Power, M. H. (1985). The effect of examiners' personal familiarity and professional experience on handicapped children's test performance. *Journal of Educational Research*, *78*, 141–146.

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: Interdisciplinary and Applied*, *114*, 37–46.

Fuchs, D., Fuchs, L. S., & Power, M. H. (1987). Effects of examiner familiarity on LD and MR students' language performance. *Remedial & Special Education*, *8*, 47–52.

Fuchs, D., Fuchs, L. S., Power, M. H., & Dailey, A. M. (1985). Bias in the assessment of handicapped children. *American Educational Research Journal*, *22*, 185–198.

Groden, J., Cautela, J., Prince, S., & Berryman, J. (1994). The impact of stress and anxiety on individuals with autism and developmental disabilities. In E. Schopler & G. Mesibov (Eds.), *Behavioral issues in autism* (pp. 178–190). New York, NY: Plenum Press.

Handleman, J. S. (1992). Assessment for curriculum planning. In D. Berkell (Ed.), *Autism: Identification, education, and treatment* (pp. 77–88). Hillsdale, NJ: Erlbaum.

Ingersoll, B. (2011). Recent advances in early identification and treatment of autism. *Current Directions in Psychological Science*, *20*, 335–339.

Johnson, C. P., Myers, S. M., & the Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, *120*, 1183–1215.

Kinnie, E. J., & Sternlof, R. E. (1971). The influence of non-intellective factors on the IQ scores of middle- and lower-class children. *Child Development*, *42*, 1989–1995.

Koegel, L. K., Koegel, R. L., & Smith, A. (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders*, *27*, 233–243.

Koegel, R. L., & Mentis, M. (1985). Motivation in childhood autism: Can they or won't they? *Journal of Child Psychology and Psychiatry*, *26*, 185–191.

Kuentzel, J. G., Hetterschedit, L. A., & Barnett, D. (2011). Testing intelligently includes double-checking Wechsler IQ scores. *Journal of Psychoeductional Assessment*, *29*, 39–46.

Marcus, L. M., Lansing, M., & Schopler, E. (1993). Assessment of children with autism and pervasive developmental disorder. In J. L. Culbertson & D. J. Willis (Eds.), *Testing young children* (pp. 319–344). Austin, TX: Pro-Ed.

Matson, J. L., Beighley, J., & Turygin, N. (2012). Autism diagnosis and screening: Factors to consider in differential diagnosis. *Research in Autism Spectrum Disorders*, *6*, 19–24.

Mesibov, G. B., & Shea, V. (n.d.). *The culture of autism: From theoretical understanding to educational practice*. Retrieved from http://www.autismuk.com/?page_id = 104

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.

Nolan, S. A., & Heinzen, T. E. (2011). *Essentials of statistics for the behavioral sciences*. New York, NY: Worth.

Olswang, 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.

Ozonoff, S., Goodlin-Jones, B. L., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 523–540.

Runco, M. A., Charlop, M. H., & Schreibman, L. (1986). The occurrence of autistic children's self-stimulation as a function of familiar versus unfamiliar stimulus conditions. *Journal of Autism and Developmental Disorders*, *16*, 31–44.

Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *Individualized assessment and treatment for autistic and developmentally disabled children: Vol. 1. Psychoeducational Profile - Revised (PEP-R)*. Austin, TX: Pro-Ed.

Stoneman, Z., & Gibson, S. (1978). Situational influences on assessment performance. *Exceptional Children*, *45*, 166–169.

Villa, S., Micheli, E., Villa, L., Pastore, V., Crippa, A., & Molteni, M. (2010). Further empirical data on the Psychoeducational Profile-Revised (PEP-R): Reliability and validation with the Vineland Adaptive Behavior Scales. *Journal of Autism and Developmental Disorders*, *40*, 334–341.

Volkmar, F., Chawarska, K., & Klin, A. (2005). Autism in infancy and early childhood. *Annual Review of Psychology*, *56*, 315–336.

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly*, *27*, 46–54.