DOCUMENT RESUME

ED 320 922 TM 015 002

AUTHOR Smith, Douglas K.; And Others
TITLE Test-Retest Reliability of the AGS Early Screening Profiles.
PUB DATE Apr 90
NOTE 15p.; Paper presented at the Annual Meeting of the National Association of School Psychologists (San Francisco, CA, April 17- 22, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Asian Americans; Black Students; Child Development; *Children; Cognitive Measurement; Comparative Testing; Correlation; Hispanic Americans; *Individual Development; Minority Groups; *Profiles; Psychomotor Skills; *Screening Tests; *Test Reliability; Thinking Skills; White Students
IDENTIFIERS American Guidance Service; *Early Screening Profiles; *Test Retest Reliability

ABSTRACT

The test-retest reliability of the American Guidance Service (AGS) Early Screening Profiles (ESP)--a battery measuring development in the cognition/language, motor, and self-help/social areas--was examined. The ESP is a nationally normed screening battery for children ages 2 years 0 months through 6 years 11 months. In addition, parent and teacher questionnaires were completed. The sample consisted of 116 children (60 females and 56 males) within the specified age range. Overall, 65% of the children were White, 31% were Black, 3% were Hispanic, and 1% was Southeast Asian. Each child was evaluated by examiners trained in administration of the battery. Retesting occurred an average of 3 weeks after initial testing, with a range of 2 days to 2 months. Pearson product moment correlations were computed; they produced the following test-retest coefficients for the immediate retest group: Cognitive/Language Profile, 0.90; Motor Profile, 0.70; and Self-Help/Social Profile, 0.81. For the delayed retest group, coefficients were: Cognitive/Language, 0.51; Motor Profile, 0.55; and Self-Help/Social Profile, 0.77. All correlations were statistically significant and fell in the moderate to strong range. The coefficients for the immediate group compare favorably with other instruments; the lower coefficients for the delayed group suggest that cognitive/language and motor skills may be very fluid at this age. Two tables summarize the data.
(Author/SLD)

Test-Retest Reliability of the AGS Early Screening Profiles

Douglas K. Smith and Michael J. Lasee
School Psychology Program
University of Wisconsin-River Falls

George M. McCloskey
American Guidance Service

Paper presented at Annual Meeting of National Association of School Psychologists, San Francisco, April 1990.

Running head: ESP

Abstract

Test-retest reliability of the AGS Early Screening Profiles (ESP), a battery measuring development in the areas of cognition/language, motor and self-help/social is examined. The ESP is a nationally normed screening battery for children ages 2 years, 0 months through 6 years, 11 months with actual testing time ranging from 15 to 30 minutes. In addition, parent and teacher questionnaires are completed in 10 to 15 minutes. The sample consisted of 116 children ages 2 years, 0 months through 6 years, 11 months (60 females and 56 males). Overall, 65% of the sample was White, 31% was Black, 3% was Hispanic and 1% was Southeast Asian. Each child was evaluated with the ESP by examiners trained in the administration of the battery. Re-testing occurred on average three weeks after the initial testing with a range of two days to two months. Pearson product moment correlations were computed and produced these test-retest coefficients for the immediate retest group: Cognitive/Language Profile = .90, Motor Profile = .70 and Self-Help/Social Profile = .81; and for the delayed retest group: Cognitive/Language Profile = .51, Motor Profile = .55 and Self-Help/Social Profile = .77. All correlations are statistically significant and fall in the moderate to strong range.

With the increased emphasis on preschool assessment, many new instruments for the assessment of preschool children's abilities and skills have been developed. An example is the AGS Early Screening Profiles (ESP; Harrison, in press), which is a nationally normed screening battery for children ages 2 years, 0 months through 6 years, 11 months. The battery measures cognitive/language, motor and self-help/social development. In addition to direct measures of skills in these areas, questionnaires are completed by parents, teachers and screening examiners. The battery produces a Cognitive/Language Profile consisting of four subtests (Verbal Concepts, Visual Discrimination, Logical Relations and Basic School Skills), a Motor Profile consisting of two subtests (Gross Motor and Fine Motor) and a Self-Help/Social Profile consisting of four domains (Communication, Daily Living Scale, Socialization and Motor). Separate scores for Expressive Language and Receptive Language Areas are determined from performance on receptive and expressive items of Verbal Concepts and Basic School Skills subtests. Standard scores with a mean of 100 and standard deviation of 15 are provided for each measure. Actual testing time ranges from 15 to 30 minutes. In addition, the parent and teacher questionnaires are completed in 10 to 15 minutes.

The Cognitive/Language subtests are administered from an easel-format. Sample items are used to communicate the task. The Visual Discrimination subtest involves the child pointing to

pictures that match stimulus pictures. In Verbal Concepts the child points to pictures of objects named or described by the examiner, and names objects pictured or described by the examiner. The Logical Relations subtest requires the child to point to pictures that correspond to stimulus pictures and to solve visual analogies. In Basic School Skills the child answers questions about number and quantity concepts, and names and recognizes numbers, letters and words.

Items on the Gross Motor subtest assess the use of legs and arms for movement and coordination, while items on the Fine Motor subtest evaluate the use of hands and fingers for manipulating objects.

The Communication Domain of the Self-Help/Social Profile assesses the child's understanding and use of oral and written language in everyday interactions. The Daily Skills Domain measures the child's ability in three categories: personal (e.g. eating, dressing); domestic (e.g. cleaning, putting away toys); and community (e.g. safety, telephone, money). The Socialization Domain evaluates the child's skills in getting along with others, playing and coping with social demands. The Motor Domain measures the child's performance of everyday fine and gross motor activities, e.g., walking, jumping. Performance in these domains is assessed through teacher/parent questionnaires.

The standardization sample for the ESP was based on 1990 census estimates and stratified on the basis of sex, race or ethnic

group, community size, region of the country and parents' level of education. The sample consisted of 1149 children with 76 to 172 children in each of 10 half-year groups between 2 years, 0 months and 6 years, 11 months.

Purpose of the Study

An important element of reliability for a new test is its test-retest reliability. Therefore, the purpose of the present study was to investigate the test-retest reliability of the ESP for two groups of children: an immediate retest group in which the retesting occurred in 21 days or less and a delayed retest group in which the retesting occurred in 22 to 60 days.

Method

Subjects

The sample consisted of 116 children ages 2 years, 0 months through 6 years, 11 months (60 females and 56 males). Overall, 65% of the sample was White, 31% was Black, 3% was Hispanic and 1% was Southeast Asian.

Procedure

Each child was evaluated with the ESP by examiners trained in the administration of the battery. Re-testing occurred on average three weeks after the initial testing with a range of two days to two months. Protocols were scored by American Guidance Service staff and double-checked for accuracy.

Results and Discussion

Mean scores for both groups of children were in the average range at both test and retest. Initial testing indicated a range of mean profile scores from 96.2 on the Cognitive Subscale Profile to 101.8 on the Self-Help/Social Profile. On retesting mean profile scores ranged from 96.2 on the Cognitive Subscale Profile to 105.7 on the Self-Help/Social Profile. Gain scores ranged from -1.1 to 3.9 for the immediate retest groups and from 0.0 to 3.6 for the delayed retest group. These results as well as test-retest reliability coefficients are presented in Table 1.

Insert Table 1 about here

Likewise, subtest scores were all in the average range for both groups at both test and retest. These scores ranged from 95.5 to 101.8 for initial testing and from 96.2 to 104.1 for retesting. Gain scores ranged from -1.0 to 2.4. These results as well as test-retest correlations are reported in Table 2.

Insert Table 2 about here

As expected test-retest reliability coefficients for the immediate retest group surpassed the coefficients for the delayed retest group. This pattern was present for both the profiles and individual subtests. Since the early childhood period is one in

which skills are rapidly developing, test-retest coefficients are likely to be lower with the passage of time. After three weeks, the test-retest coefficients are more likely to reflect stability rather than reliability.

Using r = .80 as an acceptable criterion for test-retest reliability for screening measures, the Cognitive/Language Profile and Self-Help/Social Profile, Cognitive Subscale and Language Subscale demonstrate acceptable test-retest reliability as shown by the immediate retest group. The somewhat lower reliability coefficient for the Motor Profile (r = .70) reflects the lower reliability often shown by gross motor scales, especially at younger ages.

These test-retest coefficients for the immediate retest group compare favorably with other instruments. For example, two-week test-retest reliability coefficients for the Developmental Indicators for the Assessment of Learning-Revised (DIAL-R) using a sample of 65 children were as follows: Motor = .76; Concepts = .90; Language = .77 and Total Test = .87 (Linder, 1985). Similarly, test-retest reliability coefficients on the Battelle Developmental Inventory (BDI) for 183 children over a four week interval ranged from .71 to .99 with most coefficients above .80 (Oehler-Stinnett, 1989).

The somewhat lower test-retest coefficients for the delayed retest group suggest that skills in the cognitive/language area and motor area may be very fluid at this age. Additionally, the

smaller sample size (40 versus 74) may have affected the results. Additional stability studies are needed to investigate this issue.

For both groups, Basic School Skills and Verbal Concept subtests produced the highest test-retest coefficients with the Logical Relations subtest producing the lowest (r = .69 and r = .66 for the immediate and delayed retest groups, respectively). Area scores for Expressive Language and Receptive Language were nearly identical for both groups (r = .86 to .87).

Although test-retest coefficients for individual subtests are not significantly different for the immediate and delayed retest groups, the coefficients are consistently lower for the delayed retest group. The largest differences in test-retest coefficients between the two groups are present for the Visual Discrimination (.82 versus .67) and Basic School Skills (.91 versus .78) subtests. Thus, the skills measured by these subtests may well be more fluid than the skills measured by other subtests. At the same time, the test-retest coefficients for the delayed retest groups may be affected to a greater degree by acquisition of skills and concepts than the coefficients obtained in the immediate retest condition.

Overall, test-retest reliability coefficients for language-related areas were higher than non-language areas for both groups. This result is consistent with previous research indicating verbal skills are more stable than nonverbal skills and less susceptible to practice effects. Overall gain scores in all areas and on all subtests were typically three points or less.

In conclusion, the ESP, a newly developed screening battery for children ages 2 years, 0 months through 6 years, 11 months, demonstrates adequate test-retest reliability. For the immediate retest group, reliability coefficients for profile areas ranged from .70 (Motor Profile) to .93 (Language Subscale) with a mean coefficient of .83. Total Screening Profiles produced reliability coefficients ranging from .78 (Motor + Self-Help/Social) to .89 (Cognitive/Language + Self-Help/Social). Subtest reliability coefficients ranged from .69 (Logical Relations) to .91 (Basic School Skills). A similar pattern existed with the delayed retest group with the magnitude of correlations somewhat lower.

References

Harrison, P. (in press). AGS Early Screening Profiles. Circle Pines, MN: American Guidance Service.

Linder, T. (1985). Developmental Indicators for the Assessment of Learning-Revised. In D. Keyser, & R. Sweetland (Eds.). Test critiques: Volume IV (pp. 220-224). Kansas City: Test Corporation of America.

Oehler-Stinnett, J. (1989). Battelle Developmental Inventory. In J. Conoley, & J. Kramer (Eds.), The tenth mental measurements yearbook (pp. 66-70). Lincoln, NE: The Buros Institute of Mental Measurements.

Table 1

Profile reliability data for immediate retest and delayed retest groups

	N	Standard Scores					
		Test		Retest		Gain	
		Mean	SD	Mean	SD	Score	r
Immediate Retest Group							
Profiles							
Cognitive/Language	74	98.8	13.5	101.0	14.9	2.2	.88(.90)
Motor Profile	71	100.8	14.2	99.7	13.2	-1.1	.66(.70)
Self-Help/Social	27	101.8	14.9	105.7	15.8	3.9	.81(.81)
Cognitive Subscale	74	97.7	13.5	99.4	15.1	1.7	.78(.82)
Language Subscale	74	100.1	13.2	102.2	13.3	2.1	.91(.93)
Total Screening							
Cognitive/Language + Motor + Self-Help/ Social	27	99.3	13.7	101.8	15.2	2.5	.84(.87)
Cognitive/Language + Motor	47	96.1	13.0	96.4	13.5	0.3	.82(.86)
Cognitive/Language + Self-Help/Social	27	99.6	13.7	103.3	15.9	3.7	.87(.89)
Motor + Self-Help/ Social	27	101.0	15.4	103.6	16.5	2.6	.79(.78)

Delayed Retest Group							
Profiles							
Cognitive/Language	40	96.3	11.7	97.0	14.4	0.7	.50(.51)
Motor Profile	42	100.1	14.2	103.7	13.8	3.6	.50(.55)
Self-Help/Social	34	99.5	12.5	100.4	14.0	0.6	.65(.77)
Cognitive Subscale	40	96.2	14.2	96.2	15.9	0.0	.70(.73)
Language Subscale	41	98.1	13.1	98.9	11.4	0.8	.88(.91)
Total Screening							
Cognitive/Language + Motor + Self-Help/ Social	32	100.0	14.0	101.0	14.3	1.0	.51(.53)
Cognitive/Language + Motor	40	98.3	13.8	100.0	13.4	1.7	.75(.81)
Cognitive/Language + Self-Help/Social	32	95.1	13.9	99.2	14.4	1.1	.79(.82)
Motor + Self-Help/ Social	34	100.3	14.1	103.3	14.4	3.0	.69(.73)

Note. Gain score = mean standard score from second testing minus mean from first testing.

r = Pearson correlation between standard scores. Value in parentheses is the Pearson correlation corrected for restriction of the standard score range obtained by the test-retest sample.

All correlations are significant ($p < .05$).

Table 2

Subtest reliability data for immediate retest and delayed retest groups

	N	Standard Scores					
		Test		Retest		Gain	
		Mean	SD	Mean	SD	Score	r
Immediate Retest Group							
Subtests							
Verbal Concepts	74	101.8	13.5	104.1	13.5	2.3	.83(.86)
Visual Discrimination	74	97.2	12.7	99.3	13.4	2.1	.75(.82)
Logical Relations	74	98.4	14.0	99.2	16.1	0.8	.64(.69)
Basic School Skills	74	99.1	12.4	100.3	13.0	1.2	.87(.91)
Areas							
Expressive Language	74	101.1	13.3	102.7	12.2	1.6	.82(.86)
Receptive Language	74	99.7	12.5	101.0	12.5	1.3	.81(.87)
Delayed Retest Group							
Subtests							
Verbal Concepts	42	98.2	13.5	100.6	12.0	2.4	.83(.86)
Visual Discrimination	42	97.4	13.4	97.2	15.3	-0.2	.59(.67)
Logical Relations	40	95.5	15.2	96.2	16.3	0.7	.67(.66)
Basic School Skills	41	98.2	12.2	97.2	11.3	-1.0	.67(.78)
Areas							
Expressive Language	41	99.5	12.9	100.9	12.0	1.4	.83(.87)
Receptive Language	41	98.1	12.1	99.3	11.7	1.2	.80(.87)

Note. Gain score = mean standard score from second testing minus mean from first testing.

r = Pearson correlation between standard scores. Value in parentheses is the Pearson correlation corrected for restriction of the standard score range obtained by the test-retest sample.

All correlations are significant ($p < .05$).