

ASSESSMENT OF SWIMMING PERFORMANCE OF PRESCHOOL CHILDREN¹

SARAH J. ERBAUGH

University of Wisconsin-Madison

Summary.—The purpose of this research was to establish the validity and reliability of a swimming scale designed for children, ages 2 to 6 yr. Subjects ($N = 57$) were tested on nine categories of tasks. These tasks were selected from the skills traditionally included within the motor domain of swimming; therefore, the scale is assumed to be valid. Intraclass correlation coefficients were used to estimate the interjudge objectivity, and within-day, and between-days reliabilities. The range of values for each were: .99 to .98, .99 to .96, and .97 to .84, respectively. As these values are acceptable, the swimming scale seems to be an appropriate instrument for assessing the performance of preschool children.

Even though there has been considerable interest in the development of swimming skills of preschool children, there has been limited information available on specific performance characteristics. A problem related to this dearth of information is the absence of a measuring instrument. Therefore, a rating scale was designed to assess the swimming performance of children, ages 2 to 6 yr. The purpose of this study was to establish the validity and reliability of this scale. Validity included the selection and analysis of tasks. As suggested by Safrit (1973), the validity of motor tests may be determined logically; that is, persons knowledgeable about the specific domain, in this case swimming, may identify the important performance characteristics. Reliability refers to the consistency of swimming performance as indicated by scale scores for each of the nine task categories. Three types of consistency were investigated: within-day, between-days, and between ratings of judges. Consistency or agreement between judges also may be called interjudge objectivity (Safrit, 1973).

METHOD

Subjects were 57 children (24 boys, 33 girls), ages 2 to 6 yr., enrolled in a movement program (Dec., 1975). The mean age was 4.2 yr. (Range 2.3 to 5.7) and the distribution of children by age groups was: 6 2-yr.-olds, 17 3-yr.-olds, 22 4-yr.-olds, and 12 5-yr.-olds. Children had received 10 to 50 1-hr. weekly lessons. Parents, most of whom were University faculty or graduate students, voluntarily enrolled their children.

The swimming scale has 90 tasks or items divided into nine subcategories. These subcategories represent types of swimming skills which are appropriate for preschool children. Locomotion: Front tasks, for example, may be called front crawl skills as

¹This research was conducted at the Children's Motor Development Research Laboratory directed by Marguerite A. Clifton, Purdue University. Requests for reprints should be addressed to the author at the Department of Physical Education and Dance, University of Wisconsin, Madison, Wisconsin 53706.

traditionally described by the American Red Cross. Tasks were selected for each of the categories in pilot study conducted between 1972 and 1975. Faculty very experienced in teaching children's swimming participated in the selection which included live observations and film analyses.

Tasks for each subcategory are arranged by order of difficulty—simple to complex. For example, there are 18 tasks in the first subcategory, Entry: Jump. Each task has a scale value or score. The score for the simplest task is 1 and the score for the most difficult task is 18. The child who is lifted into the pool by the examiner is given a score of 1, and a child who climbs into the pool with the assistance of the examiner is given a score of 2. Finally, the child who jumps independently into the water with depth of 6 ft.; touches the bottom; and then propels self, using an underwater pattern, a distance of 10 ft. is given a score of 18. In this example each task represents a recognizable step along the Entry: Jump continuum. An advantage is that scores for an individual child may be interpreted by referring to the original scale values, and a limitation is that the distance between each task is arbitrary, and it is not assumed to be equal across tasks (Wohlwill, 1973).

The general descriptions of performance and the number of tasks in each subcategory are:

1. Entry: Jump: Enter pool foot first by climbing down ladder or jumping into water and propelling self to examiner; 18 tasks.
2. Locomotion: Front: Maintain prone position and propel self with or without assistance of examiner or flotation device; 14 tasks.
3. Locomotion: Back: Maintain supine position and propel self with or without assistance of examiner or flotation device; 15 tasks.
4. Breathing: Submerge face or all of head and exhale or blow bubbles for several seconds; 15 tasks.
5. Kicking: Maintain prone position and propel self using legs only, with or without assistance of examiner or flotation device; 14 tasks.
6. Diving: Enter pool head first and propel self to teacher; 3 tasks.
7. Ring Pick-up: Stand independently in shallow (chest deep) water and grasp ring placed near feet; 4 tasks.
8. Ring Retrieval: Surface dive in deep water (4 ft.) and grasp ring placed on bottom of pool; 3 tasks.
9. Hoop Obstacle Course: Maneuver or swim through several hoops positioned at water surface level; 4 tasks.²

Two judges and an adult swimmer (examiner) were required for test administration. Judges, one of whom was the investigator, were faculty members and examiners were female undergraduate students. All testing personnel were certified Water Safety Instructors with previous experience in teaching preschool children to swim. Personnel attempted to keep testing conditions similar for each child; that is, the directions of the examiner, the arrangement and use of the equipment, and the order of performance of tasks were similar.²

Children were tested individually while at least one parent observed from the pool deck. The time required for completion was approximately 20 min. which included 5 min. of practice. During the test, the examiner asked the child to perform two trials for each subcategory of tasks, and the judges independently rated the performance. One week following the original test a random subsample of 22 children was retested.

²Lists of tasks for subcategories and descriptions of examiner's procedures are available from the author.

The pool dimensions were 60 ft. \times 30 ft. and the depth was 4 ft. to 6 ft. This depth was appropriate for six of the task subcategories (1, 2, 3, 5, 6, 8). Three subcategories required a water depth of 2 ft., 6 in. (4, 7, 9). Two pieces of equipment were used: first, a 6-ft. \times 6-ft. platform for Breathing and Ring Pick-up tasks, and second, a 12-ft. \times $\frac{3}{4}$ -ft. bench for Hoop Obstacle Course tasks. Also, lane lines were used by the judges to determine distance. The temperatures of the water and the room were 86° and 90°, respectively.

RESULTS AND DISCUSSION

Data for interjudge objectivity, intertrial reliability, and test-retest reliability are presented in Table 1. Intraclass correlation coefficients (r) were calculated using analysis of variance procedures. Also, percentages of agreement between judges were calculated for interjudge objectivity (Safrit, 1976; Baumgartner & Jackson, 1975; Werner & Bayley, 1966). Means and standard deviations of the swimming performance scores for the sample ($N = 57$) and the subsample ($N = 22$) are presented in Table 2. Preliminary data analysis deemed the selected parametric procedures appropriate.

Data used for interjudge objectivity were the scores of two judges for performance of sample on the second trial. Intraclass correlation coefficients ranged from .99 to .98. These high values indicated that the two judges were very consistent. Percentage of agreement ranged from 96.4% to 85.9%. Low values such as 85.9% for Breathing scores indicated that judges sometimes disagreed. One explanation for the disagreements is that the performance of preschool children is extremely variable especially on Breathing tasks.

Data used for intertrial or within-day reliability were the performance scores of sample for two trials as reported by one judge, the investigator. Intraclass correlation coefficients ranged from .992 for Locomotion: Back to .969

TABLE 1
RELIABILITY FOR SWIMMING PERFORMANCE SCORES OF PRESCHOOL CHILDREN

Task Subcategories	Interjudge Objectivity*		Intertrial Reliability*		Test-retest Reliability**
	Intra-class r_s	% Agreement	Intra-class Correlations r (2 trials)	r (1 trial)	Intra-class r_s
1. Entry: Jump	.99	92.9	.985†	.970†	.94
2. Locomotion: Front	.99	87.7	.982	.965	.97‡
3. Locomotion: Back	.99	94.7	.992	.983	.95
4. Breathing	.99	85.9	.986†	.986†	.84
5. Kicking	.99	92.9	.973	.947	.92
6. Diving	.99	96.4	.991	.982	.94
7. Ring Pick-up	.98	94.7	.969	.941	.90
8. Ring Retrieval	.98	94.7	.983†	.967†	.97
9. Hoop Obstacle Course	.99	96.4	.976	.954	.89

* $N = 57$. ** $N = 22$.

†Trial effect was significant, therefore, variance due to trials was not included in the error term.

‡Day effect was significant, therefore, variance due to days was not included in the error term.

TABLE 2
MEANS AND STANDARD DEVIATIONS FOR SWIMMING PERFORMANCE
SCORES OF SAMPLE AND SUBSAMPLE

Task Subcategories	Sample (N = 57)		Subsample (N = 22)			
	M	SD	Test		Retest	
			M	SD	M	SD
1. Entry: Jump	9.93	5.09	11.13	4.62	10.84	5.06
2. Locomotion: Front	7.03	3.15	7.72	3.28	7.13	3.91
3. Locomotion: Back	6.01	4.34	7.31	4.97	7.40	4.42
4. Breathing	8.31	3.88	9.68	3.60	9.36	3.24
5. Kicking	8.80	3.60	9.45	3.39	9.00	3.55
6. Diving	0.82	1.02	1.18	1.14	1.13	1.24
7. Ring Pick-up	1.64	1.47	2.32	1.46	2.50	1.37
8. Ring Retrieval	0.68	1.02	0.95	1.13	1.09	1.19
9. Hoop Obstacle Course	1.89	1.31	2.36	1.09	2.50	1.22

for Ring Pick-up. These values indicated that children performed very consistently across trials. As fatigue was often encountered with two trials for each subcategory, reliability estimates were computed for one trial. These values ranged from .986 for Breathing to .941 for Ring Pick-up. The comparison of correlations based on two trials with one trial, showed very small differences. These data suggest that one trial is almost as reliable as is the mean of two trials.

Data used for test-retest or between-days reliability were the second trial scores of subsample as reported by one judge, the investigator. Intraclass correlations ranged from .97 to .84. Correlations were $\geq .90$ for seven of the task subcategories, and the highest value of .97 was reported for Locomotion: Front. Correlations were $< .90$ for two of the task subcategories, and the lowest value of .84 was reported for Breathing. Thus children were most consistent on Locomotion: Front tasks, and they were least consistent on Breathing tasks. Also, test and retest means in Table 2 indicated very little change in performance between days.

In summary, the reliabilities of the swimming scale are acceptable; therefore, the instrument may be used with confidence. Intraclass correlations for intertrial consistency were $\geq .96$ and for between-days consistency were $\geq .84$. Those for between-judge consistency were $\geq .98$, and percentages of agreement were ≥ 85 . In addition, the instrument is assumed to be valid because of design of tasks.

REFERENCES

- BAUMGARTNER, T. A., & JACKSON, A. S. *Measurement for evaluation in physical education*. Boston: Houghton Mifflin, 1975.
- SAFRIT, M. J. *Evaluation in physical education: assessing motor behavior*. Englewood Cliffs, N. J.: Prentice-Hall, 1973.
- SAFRIT, M. J. (Ed.) *Reliability theory*. Washington, D. C.: American Association for Health, Physical Education, and Recreation, 1976.
- WERNER, E. E., & BAYLEY, N. The reliability of Bayley's Revised Scale of Mental and Motor Development during the first year of life. *Child Development*, 1966, 37, 39-50.
- WOHLWILL, J. F. *The study of behavioral development*. New York: Academic Press, 1973.

Accepted November 21, 1978.