A GRADED SERIES OF GEOMETRICAL PUZZLES

BY GRACE HELEN KENT

This series of puzzles is offered as one unit of a group of non-verbal tests to be used for measuring the capabilities of defective children. The work was commenced in the School of Pedagogy of New York University. The general trend of the undertaking has been influenced by three months' experience in the use of the Binet-Simon measuring scale (Goddard's translation) in testing children of the ungraded classes of the public schools of New York and Jersey City. The majority of the children examined were high-grade defectives, but the group included some middle- and low-grade imbeciles and also a few mentally normal children who had been placed in the special classes because of nervous disorders. Nearly all were children of foreign-born parents.

My experience in the use of the Binet test is too limited to justify a general criticism of the method, but I find it necessary to mention four fundamental difficulties which I encountered: (1) The scoring of results is arbitrary, and depends too much upon the personal equation of the examiner. (2) There is little or no uniformity in the tests for children of different levels of intelligence. (3) The test depends too much upon the subject's command of the language in which it is given to be valid for children of immigrants. (4) The subject matter of the test is too far removed from the interests of children to hold their attention, and it frequently requires much urging to induce them to cooperate.

Yerkes\(^1\) has called attention to the lack of uniformity and the inaccuracy of the system of scoring results, and has indicated a way of overcoming these difficulties. He has also shown that the disadvantage which the test imposes upon children of non-English-speaking parents is smaller than

\(^{1}\) 'The Point Scale: a New Method for Measuring Mental Capacity,' Boston Medical and Surgical Journal, December 3, 1914.
might be expected and that it is comparatively constant for children of different ages. In testing children of immigrants by groups, it might be possible to make allowance in the final scoring of results for this disadvantage; but if children are to be tested as individual cases this would not be a safe plan to follow, inasmuch as the deficiency in the language might be so marked as to entirely invalidate the results for the particular child in question.

The fourth objection, the problem of interest, is in my opinion the most vital of all. The purpose of the test is not to determine the child’s willingness to answer certain questions, but to determine his ability to answer them. If his cooperation is given unwillingly, the results are misleading. It is plain that Binet recognized this factor, for he called attention to the need of leading the child to regard the test as a game. But this does not depend wholly upon the attitude of the examiner. Some children, especially high-grade defectives, are extremely sensitive to any reminder of their deficiencies. Such a child may be irritated by a question that he cannot answer, and he is liable to feel humiliated when required to answer one that is entirely too easy for him. As it is impossible to determine his level without asking at least a few questions above and below it, the test is something of an ordeal. Young children are sometimes painfully shy in the presence of a stranger, and it requires no small effort to induce them to speak a single word. Such a child may be pleased when a picture is shown, but he becomes self-conscious as soon as he is requested to describe it. He may name a single object, then wait for the question “And what else do you see?” before he ventures to name another. It is possible that a child’s willingness or unwillingness to talk to a stranger may be in itself of some significance as an index to his intelligence; but the Binet test is inadequate as a means of determining the exact degree of communicativeness, while it is of course an unnecessarily elaborate method of ascertaining that a child is or is not talkative.

Even the highly practised observer finds it difficult to give scrupulous cooperation in a psychological experiment that
involves distasteful tasks. And in order to obtain full cooperation from immature, defective, insane, or otherwise irresponsible subjects, it is important to offer something that will stimulate a genuine desire to make a good record.

These considerations have led me to undertake the development of a system of non-verbal tests. In attempting this I do not wish to discount the importance of any verbal tests that have been found serviceable in the diagnosis and classification of defectives. In general these tests of action are intended to supplement rather than supplant such verbal tests as are fairly applicable to the subject. It would seem that they should be especially helpful in the study of children of foreign parentage, and possibly in the examination of immigrants at Ellis Island. In the selection of test methods I have adopted provisionally the following rules:

1. Each test is to call for a motor reaction from the subject, rather than for a verbal response. The test itself is to be essentially mental, not motor; but it is to be a test of the subject's ability to perform a certain act which calls for some mental effort, not his ability to answer certain questions.

2. Each test is to require the minimum of verbal explanation on the part of the examiner, so that the subject's attention will not be seriously taxed in merely comprehending the requirements. The instructions are to be so simple that they may be given satisfactorily through an interpreter.

3. Each test must possess sufficient intrinsic interest for the subject to hold his attention and to command his spontaneous cooperation. The performance of the act must possess a certain dignity, so that the subject will not consider it an insult to his intelligence. The accomplishment of the task is to be, from the subject's point of view, an end in itself. The aim is not merely to devise tasks that can be presented attractively by a kindergartener; it is rather to offer something which is in its very nature a game, and so good a game as not to suffer total shipwreck in the hands of the ultra-scientific examiner.

4. Each task is to be capable of being graded, so as to be adaptable to the ability of any subject. The method
must admit of being varied widely in difficulty without affecting the essential nature of the test.

5. Each task must admit of being evaluated by a system that will give partial credit for partial success. The results are to be scored either by a sliding scale or by empirically determined steps.

The unit here presented is well adapted for use among deaf-mutes or foreigners having no knowledge of the language of the examiner, because the instructions may be given wholly by gesture. There are, of course, many similar puzzles already in use; but in view of the present demand for graded tests, it seems probable that this set of puzzles will meet a real need.

The series includes twenty-four puzzles, consisting of rectangular or triangular pieces of wood. For convenience I have arranged the rectangular and the triangular puzzles independently, in parallel series (see plate). Each puzzle when solved forms a square, and is solved by fitting the pieces into a frame ten centimeters square. The puzzles were cut from wood 3/16 inch in thickness. The frame was made by attaching strips of wood 1/16 inch in thickness to the margins of a square of wood, leaving a depression ten centimeters square. The absolute thickness of the wood is presumably of no importance, but it is important that the puzzle pieces should be thicker than the border of the frame, so that any piece which has been placed in the frame may still be easily accessible to the fingers. If desired, cardboard may be used instead of wood. A full set of puzzles, including the frame, may be made very easily by pasting together several thicknesses of light cardboard that can be cut with scissors.

In the presentation of the puzzles to the subject it is well to alternate between rectangles and triangles, because this tends to reduce the effect of practice. The time occupied by the solution of each puzzle is measured by a stop watch. In order to prevent the test from consuming more time than can reasonably be allowed for it, and also to keep a child from becoming unduly fatigued by the effort to solve a puzzle
that is hopelessly beyond him, I have found it necessary to set an arbitrary time limit. When the gradation of the puzzles has been more thoroughly tested, it may be possible to fix the time limit for each puzzle according to its difficulty. For the present I allow two minutes for each puzzle, and at the close of that period I interrupt the subject and record that attempt as a failure.

The four puzzles consisting of less than three pieces are designed primarily for infants or subjects of very low-grade
intelligence, and these need not be given to a subject of fair ability. In a routine test I usually begin with 3, 4, or III., and give the puzzles as nearly in the order of difficulty as the alternation between rectangles and triangles permits, until the subject has failed to solve (within the time limit) at least one puzzle of each kind.

The arrangement of the puzzles is somewhat tentative, being based upon results obtained from but twenty-five subjects. In the development of the series preliminary tests have been made upon over one hundred subjects, including sixty-nine defective children. But the results of these tests were obtained under somewhat irregular conditions, and it seems hardly advisable to attempt to make them presentable. I have therefore disregarded them and have made a special series of tests upon twenty-five normal subjects, using the first twenty-five persons who were by any chance available. They range in age from six to seventy-five, and in education from illiteracy to university training.

The results of this series of tests are shown in tabular form. As far as possible, the arrangement of the puzzles is based on the median time record of the twenty-five subjects. The puzzles which were solved within the two-minutes time limit by less than half of the subjects are arranged according to the number of failures which they elicited.

In the early preliminary tests I used no frame, but requested the subject to make a square of the pieces. But the results were very variable, and some of the younger children failed to understand what was meant by a square. The use of the frame is recommended for various reasons, especially because it lowers the scale as a whole. In the grading of a mechanical test it is the lower end of the scale that presents the greatest difficulty. It is well to make the test qualitatively as simple as possible, and then to bring it to any desired degree of difficulty by quantitative differences.

Other things being equal, the difficulty of a puzzle depends upon the number of pieces. But a puzzle made of triangles is far more difficult than one consisting of the same number of rectangles, and there are indications that the two types differ in kind as well as in degree. The correlation betwe
### Table I

<table>
<thead>
<tr>
<th>Rectangular Puzzles</th>
<th>Triangular Puzzles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Puzzle</td>
<td>Minimal Time Record in Seconds</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>—</td>
</tr>
<tr>
<td>14</td>
<td>—</td>
</tr>
</tbody>
</table>

them is apparently rather weak, and it is an open question whether they should be treated as one test or as two tests.

Puzzles 8, 9 and II. are somewhat superfluous, and need hardly be considered as belonging to the series. They are included on the ground that it is frequently convenient to have an extra puzzle which may be used as a control test. In case a child solves puzzle I. very promptly, apparently by chance, it is well to give puzzle II. Puzzles 8 and 9 consist of the same number of pieces respectively as IO and II, but the former are simpler of construction. If a subject is able to solve IO and II, it is a waste of time to require him first to solve 8 and 9. But if he fails to solve IO, it is only fair to permit him to try the intervening puzzles, so as to insure that his limit has been reached.

In the preliminary tests thirty-eight puzzles have been
tried, and those which yielded the most variable results have been gradually weeded out. The figures $x$ and $y$ represent two puzzles that proved to be exceptionally unsatisfactory. Both gave a very low minimal time record but a very high percentage of failures, as compared with other puzzles of approximately similar difficulty. This indicates that the element of chance is an important factor in the solution. Puzzle $x$ is constructed upon a definite plan, a plan which the subject may grasp at sight, or which he may fail to grasp after several minutes of hard work. Puzzle $y$ is apparently very simple, but many subjects are misled by the not unnatural assumption that a right angle properly belongs in a corner of the frame. These two puzzles, however, are useful when there is occasion to give the instructions by gesture. As they are quite unlike any puzzles of the series, they may be used freely by way of illustration.

The puzzles which yield the most uniform results are those which call for a patient and systematic use of the trial and error method. Of course the element of chance cannot be eliminated, but it can be greatly reduced by avoiding any obvious principle of construction. In the rectangular puzzles the pieces should vary in both dimensions, and in the triangular puzzles it is well to depend mainly upon oblique or scalene triangles.

The puzzles at the lower end of the scale have seen considerable use in the tests made upon defective children, and it is significant that puzzles I and 2 are the only ones which have met with universal success. I have observed several children work very patiently for the full two minutes with puzzles I. and II., without success. It would seem that a two-piece puzzle must certainly be solved by chance within the time limit, but there are exceptions to this. These puzzles should be tested by using them with normal infants. I have as yet had access to only two infant subjects, one of twenty-three months and one of twenty months. The older child was successful with puzzles I and 2, but could not be induced to give sufficient attention to any others. The younger child was successful only with puzzle I, which is nothing but a form board. Of course the failure to gain the
coöperation of so young a child must be scored against the resources of the examiner in arousing the interest of the child, rather than against the child's ability.

I have not determined at what age a child may be expected to show interest in the puzzles, but I have found it very well developed in two children of six years. Older children are fascinated by them, and it is surprising to see a healthy boy so absorbed in any sedentary occupation. Some of my adult subjects have manifested an interest almost equal to that of children.

Even the defective children showed considerable spontaneity in their reactions to the puzzle test, and I made a practice of interspersing the puzzles among the questions of the Binet test, so as to make the test period as a whole more attractive. I have rarely found a child who would voluntarily abandon the attempt to solve a puzzle. Most children are reluctant to give up when time is called, and I have frequently been moved to promise another trial at the close of the test. A boy of twelve once remained more than an hour, working at the puzzles with absorbing interest and taking little notice of three persons who were watching him. His efforts were remarkably unintelligent, and there was no observable tendency to profit by experience. Because of his great persistency he finally succeeded in solving several of the moderately difficult puzzles; but if requested to try again one which he had just solved, he was as helpless as before.

The subject's mode of reaction is frequently more instructive than the time record, and it should always be observed. I have not made a practice of recording false moves, because I have not found any satisfactory way of classifying them; but I record under 'Remarks' anything noteworthy in the subject’s method of working.

The puzzles should be kept behind a screen, and the subject should not be allowed to know how many there are. It is well to explain in advance that nobody is expected to be successful with every puzzle, so that he will not be disconcerted by his first failure. If he drops out in the early part of the series, he should be spared the annoyance of knowing how much he is leaving undone.
The significance of this test as a means of measuring intelligence remains to be seen. But inasmuch as geometrical puzzles are actually being used for clinical purposes, it seems worth while to develop a more systematic method of using them, so as to determine their possibilities and their limitations. The ability to grasp the relation between area and form is a highly specialized function, and its correlation with general intelligence may be comparatively weak. One of my subjects, a man of national reputation in his own field, spent nearly three minutes solving a puzzle which has since been solved in twenty-seven seconds by a child of six years. This puzzle (figure x) has since been dropped from the series, but the fact remains that subjects of unquestioned intelligence vary very widely in their possession of the ability to solve puzzles of this type.

If mechanical tests are to be useful for determining intelligence, it is of the utmost importance to have available more and yet more tasks which differ essentially in kind and which call into play widely different types of mental activity. Among the units of such a group there should be pairs which show a strong negative correlation, and the tests should be so varied as to give any subject a fair opportunity to show what kind of ability he may possess.

There will be no attempt at present to standardize this test on a large scale, because it is not certain that it is worthy of being standardized. The present need is for intensive rather than extensive study of test methods. A large number of methods must be tried out in order to find even a few that will be successful, and it is a work of years to find a sufficient number and variety of tasks each of which shall meet all the practical requirements. Standardization, of course, does not render valid a method that is fundamentally defective, but it is all too likely to prolong the life of a criterion which would far better be permitted to die a natural death.

I am offering this unit for what it is worth, possibly a little prematurely, in the hope that its use by others will bring to light its possibilities for improvement. Criticism both of the general plan and of this particular unit will be greatly appreciated.