Irish Journal of Education, 1971, v, 2, pp 59-69

ISSUES IN THE MEASUREMENT OF INTELLIGENCE

WILLIAM M BART
University of Minnesota

A variety of problems arising out of current practices in the measurement of intelligence are considered These include the gross imprecision of definitions of intelligence and the norm referred basis of intelligence-test construction It is argued, primarily from a Piagetian framework, that the precise measurement of intelligence requires the construction of custom-made structurally parallel intelligence tests If such tests are also criterion-referenced, the information provided to teachers and counselors regarding the intelligent behaviours of students will be much more useful than the information available on the basis of performance on tests currently in use

THE NEED TO DEFINE INTELLIGENCE

A crucial problem in the measurement of intelligence relates to the paucity of precision in many definitions of the term Some intelligence tests have been constructed from reflections on such nebulous definitions as 'innate general ability' (1) and 'what intelligence tests measure' (27) The content validity of these tests must be held suspect since the underlying definitions of intelligence are not explicit in their designation of constituent behaviours

Some theorists, it is true, have attempted to provide fairly specific criteria of intelligent behaviour Charles Spearman (21), for example, contended that intelligence or *g* (a general intellective factor) relates to the eduction of either relations or correlates Thus, test items which require either the eduction of relations (e g , similarities) or the eduction of correlates (e g , analogies) may be regarded as measuring intelligence Though Spearman's view of intelligence may be somewhat narrow, it does permit one to select behaviours which are manifestations of intelligence Theorists, such as Thurstone, Vernon, and Burt, have made extensive use of factor analytic techniques to examine the nature of intelligence They have tended to claim that a given item measures a particular factor in intelligence if the item loads highly on the factor under consideration This, too, is rather narrow and circular While Guilford (8) supports the psychometric criterion of intelligence, he also makes use of behavioural criteria In his view, there are in excess of seventy separate abilities constituting intelligence This set includes the two abilities described by Spearman as well as other ones such as the ability to cognize semantic units, which is tested with vocabulary definition items The set of abilities or factors listed by Guilford also provides some basis for distinguishing intelligent from non-intelligent behaviours

The psychometric tradition on the whole has not been overly concerned with defining intelligent behaviour As Margaret Donaldson (3) has pointed out, test constructors rely to a large extent on hunches in the selection of items and generally cannot justify their choice on theoretical grounds It would seem important, however, in measuring intelligence to define the concept clearly so that it is possible to decide whether or not some human behaviour is an example of intelligence In other words, what is included in the definition of intelligence must be sharply differentiated from what is not included It is only when intelligence has been well-defined that content validity and sampling validity (which indicate the extent to which a test is actually measuring the behaviour defined) can be satisfactorily established (9) Precise definition does not preclude the possibility of a variety of definitions But it is only when intelligence has been well-defined that one can determine which behaviours are common to any two definitions and which are unique to a particular definition Factor analytic procedures, such as those cited by Joreskog (13), could be used for this purpose

The view that there is a need for improved definitions of intelligence has its critics Jensen (12) has even argued that there is no reasonable answer to the question 'what is intelligence?' Intelligence, he says, is already measurable and with quite substantial success He also indicates that some things like intelligence and electricity can be more readily measured than defined The view that intelligence is similar to electricity is somewhat misleading since electricity complies to such specific mathematical laws as those of Ohm, whereas there are no specific mathematical laws, free from sample-based statistical elements, relating intelligence to any other entity Since there are invariant laws of electricity, electricity can be regarded as being well-defined, an entity is considered to be electricity only if it conforms to the laws of electricity No such well-defined status presently exists for the construct of intelligence Until it does, the entity as presently measured is of questionable value in determining the structural basis of the science from which it is drawn

This view leads one to challenge another point of Jensen—namely, that intelligence is quite easily and successfully measured. Jensen's view is based on two premises first of all, that there is a substantial general factor underlying intelligence tests, and secondly, that there are substantial predictive correlations between intelligence test scores and scholastic achievement measures The success of the measurement of intelligence, however, lies primarily in the fact that certain psychometric properties are manifested rather than that a high degree of either content or construct validity is indicated Though intelligence may have been easier to measure than to define, that does not obviate the need to look for precise definitions of the term and for scientific laws relating intelligence to other psychological variables

ORDER OF DIFFICULTY OF MEASURES OF INTELLIGENCE

Another problem in the measurement of intelligence is the determination of the order of difficulty of tasks or test items designed to measure specific intelligent behaviours In this context, an intelligence test composed of twenty items can be considered as being made up of twenty measures of intelligence Within a Piagetian framework of intelligence measurement, the plasticine ball tasks measure for the concept of conservation of mass, a physical task measures for the concept of conservation of weight, and the equilibrium in the balance task measures for the schema of proportionality These measures comply to a linear order of difficulty with the ball task being the easiest and the balance task the most difficult. The order of difficulty for these measures is considered to be universally invariant—i e , the linear order of difficulty for the three measures should be the same for any sample of human subjects Margaret Mead, in a work edited by Tanner and Inhelder (22), has indicated that the order in which Piagetian stages of intelligence appear is probably constant in any culture known to her, and there is some empirical evidence confirming the cross-cultural quality of the invariant order of Piagetian behaviours Price-Williams (20), for example, determined that non-westernized African children pass through the three stages of conservation (both for continuous and discontinuous quantities) in the same order as European and other Western children Other measures of intelligence may not engender such an extensive level of invariance in their order of difficulty, that is, the order may be invariant only within certain cultural groups The determination of the orders of difficulty of measures of intelligence should be given high priority in the development of intelligence tests

Another way of viewing the problem of order of difficulty in measures of intelligence (i e , individual intelligence items or tasks) is to consider the logical relationship between success on one measure and success on another measure for any array of measures That is, it should be possible to specify that success on one specific task is a necessary but not sufficient condition for success on another specific task The logical relationships between tasks may take on the forms of implication and equivalence Once sufficient research has been carried out on the logical relationships between measures of intelligence, then it will be possible to construct a battery from a subset of the measures, the selected measures may be required to form a linear hierarchy so as to allow the formation of an ordinal scale

Intelligence may be described as a repertoire of ways of structuring entities As one grows, one's intelligence grows, for the number of structurings available to the individual will most likely increase The concept of intelligence as a set of fundamental cognitive 'tools' with which problems are resolved emanates from the concept of intelligence of Jean Piaget (16) An example of such a way of structuring, or intellectual tool, is seriation by which an individual is able to order entities according to some dimension Many of the constituent elements of intelligence (ways of structuring such as seriation and classification) form a linear hierarchy (16) Using this framework, the individual with the greater repertoire of ways of structuring the environment is regarded as the more intelligent A level of intellectual functioning may be defined in terms of ways of structuring the environment that develop and are manifest concurrently

The development of intelligence may then be described in terms of an ordinal scale of levels of intellectual functioning Each level of intellectual functioning should be defined in terms of a class of behavioural events relating to ways of structuring so that a behavioural event at one level of intellectual functioning must either imply or be implied by a behavioural event at another level An example of a level of intellectual functioning is concrete operational thought which is characterized by such activities as seriation and classification of objects With this manner of definition, an index of intelligence would refer to level of intellectual functioning and not, as does the IQ index, to the position of an individual in a given population

Intelligence tests constructed in harmony with an exacting well-defined definition of intelligence incorporating levels of intellectual functioning would be criterion-referenced rather than norm-referenced (7) Criterion-referenced tests would be diagnostic as they would provide explicit information on the ways of structuring the environment (or the intellectual skills) available to an individual as well as about those not available They would also be composed of items that require for their correct solution the use of the intellectual skills enumerated in the definition of intelligence

In the construction of most intelligence tests now in use, a pool of presumably content-valid items is first developed and from this pool the actual test items are selected on the basis of psychometric properties such as difficulty and discrimination indices Further, these tests are norm-referenced since they provide information as to the placement of the testee with respect to a normative population Criterion-referenced intelligence tests on the other hand would provide information on the placement of testees on a scale of intellectual development, how intelligent a testee is relative to other testees would be of secondary importance

An example of a criterion-referenced intelligence test item, based on a Piagetian conception of intelligence, is the Equilibrium in the Balance task (11) In that task, an individual, when confronted with a balance with weights, is asked questions by an interviewer to determine whether or not he is aware of the law of the balance which incorporates the schema of proportionality, which is a constituent process of formal reasoning If the individual successfully resolves the balance task, he is considered to be capable of formal thought which is the highest level of intellectual functioning in the Piagetian system, if he fails, he is considered to be at a lower level of intellectual functioning With tasks and items such as the balance task, criterion-referenced intelligence tests could be constructed that would indicate which intellectual capabilities a subject has and which ones he does not have Also, items could be formulated to make finer discriminations within levels of intellectual functioning For example, according to Inhelder and Piaget (11) there are levels of combinatorial reasoning within the stage of formal operations Each level is determined by the number of bivalent variables in a problematic situation for which a subject can generate all of the possible relationships It should be possible to determine combinatorial reasoning levels by varying the number of variables to be examined in test items

THE MANIFESTATION OF INTELLIGENCE

A third problem in the measurement of intelligence relates to a misconception regarding how intelligence manifests itself There is an implicit belief by many theorists (e g , Spearman, Jensen) that the measured level of intelligence for an individual in a certain test context would be manifest in a great variety of contexts, and in the construction of intelligence tests there has been little concern for possible item context effects on manifested levels of intelligence The author contends, as Flavell and Wohlwill (6) do, that the item context effect may be sizeable in increasing the difficulty of an intelligence test item and that the level of intelligent behaviours manifested in any test situation is inextricably tied to the context of the situation A person may manifest a higher mode of intellectual functioning in one context than in another Two items testing for the same intelligent behaviour may have different contexts with differences along such dimensions as content area (e g , mathematics, history) and modality (e g , written, oral) Sir Cyril Burt (2) made the same point when he pointed out that group tests of verbal intelligence are inappropriate for assessing intelligence in the case of children from disadvantaged environments, he contended that non-verbal, open-ended intelligence test items need to be constructed to test children whose abilities have developed along non-conventional but inventive lines

In measuring intelligence, it is common to attempt to determine the highest level of intellectual functioning available to an individual, in other words, the 'level of intelligence' of an individual is the optimal level of intelligence available to him (10) It follows that the level of intelligence determined for any given individual should always be viewed as being a conservative estimate since it is always possible that the individual may manifest a higher level of intellectual functioning in some untested context It is hypothesized that an individual will manifest his highest level of intellectual functioning in a context familiar to him One empirical study lends support to this view Kellaghan (14) tested westernized and non-westernized Yoruba children, aged about eleven years, for the ability to abstract using three tests the Goldstein-Scheerer cube test, the Weigel-Goldstein-Scheerer colour-form test, and an object-sorting test made up of objects equally familiar to both groups No significant difference between the performance of the two groups was found on the object-sorting test in which the content was familiar to both However, significant differences in favour of the westernized Yoruba were found on the other two tests, presumably because the items employed form concepts and other Occidental patterns which were familiar to the westernized children but were not familiar to the non-westernized ones Thus, with an African people it is confirmed that children tend to manifest higher levels of intelligence in contexts familiar to them

Another purpose in the measurement of intelligence is to determine the lower limit of the level of intelligence of a testee Tests should therefore attempt to minimize the underestimation of intelligence The results of intelligence tests are often used to help teachers and counsellors make decisions about a student's capability to follow a particular curriculum or course If intelligence is underestimated, as is often the case with minority group students, many educational activities and opportunities may be closed off to those students who are judged not to have the required cognitive capabilities A student whose intelligence is underestimated may, for example, be coerced to enter a vocational programme of instruction and to become a tradesman while, if given the opportunity, he might have greatly enjoyed and found intellectually stimulating a more academic type of programme Thus, because of the awesome social ramifications of the use of intelligence tests (e g , prediction of academic success, classification of students), it is crucial that intelligence be neither underestimated nor overestimated

With respect to the predictive role of intelligence tests, some evidence has been collected to indicate that non-conventional measures of intelligence can be better predictors than conventional ones Dudek *et al* (4) determined that a battery of nine Piagetian tasks drawn from the Montreal scale of intelligence was a superior predictor of first-grade achievement test scores than either the Wechsler Intelligence Scale for Children or the Lorge-Thorndike Group Intelligence Scale However, possibly more crucial than the predictive capabilities of non-conventional intelligence tests are the diagnostic and prescriptive capabilities of such tests With structurally-parallel intelligence tests that are criterion-referenced, the intellectual abilities and inabilities of a pupil for each of a variety of content areas could be determined It is possible that curricula may be altered, providing more familiar content for students who are unable to manifest their intellectual abilities in conventional content Also, more research may be carried out to determine the most efficient methods of effecting transfer or *horizontal decalage* across various content areas for specific intellectual skills The findings of this research would, hopefully, provide guidelines for helping children from various cultural backgrounds to learn to apply efficiently the intellectual skills they have to the mastery of new knowledge areas in the formal educational setting Intelligence tests providing accurate estimation for a group of subjects must be custom-made Such tests would employ contexts confirmed to be within the range of experience of the subjects Thus, a set of structurally-parallel intelligence tests set in various contexts may have to be used to determine the lower limits of intelligence of individuals coming from different backgrounds Such tests would be composed of structurally-parallel items in which equivalent intelligent behaviours are required for the correct resolution of corresponding items and in which different contexts are also used for corresponding items

Information on the generalizability of the intelligent behaviours available to an individual should also be provided to indicate to teachers and counsellors the range of contexts in which an individual can employ certain modes of reasoning This information would be helpful in planning an academic programme for an individual, for if one can employ higher modes of reasoning in one context than in another, then instruction for the individual in the former context may utilize higher modes of reasoning than instruction in the other context With the use of a battery of structurally-parallel intelligence tests, it would be possible to determine a profile of an individual's intelligence, such a profile would indicate the generalizability of the intelligent behaviours available to an individual as well as the level of intellectual functioning proper to the individual for each of several contexts The best estimate of intelligence for an individual would be the greatest level of intelligence recorded for him on the tests

Practically all modes of measurement may be represented in structurally-parallel intelligence tests (paper-and-pencil instruments, observation techniques, physical tasks) In fact, a variety of contexts with various modes of measurement will probably be necessary if one is to determine accurately the level of intelligence of an individual For example, a paper-and-pencil intelligence test may reveal a low level of intellectual functioning for a Laplander living in a large city, however, a set of physical tasks employing materials familiar to the individual, and constituting a structurally-parallel intelligence test, may indicate a higher level of intellectual functioning With such precise culturally appropriate tests there is substantial likelihood that intergroup and intercultural differences in intelligence test performance will be markedly reduced However, the educational value of testing will be maintained, since the tests will be helpful in deciding whether the individual's difficulty with a certain content area reflects his inability to transfer available modes of reasoning into the context of the content area or whether it reflects a lack of development of modes of reasoning required by the test A lack of ability to transfer intelligent behaviours and a lack of the development of such intelligent behaviour call for two quite different courses of educational action Thus, structurally-parallel custom-made intelligence tests could provide much valuable information to educators

THE DEVELOPMENT OF NEW MEASURES OF INTELLIGENCE

There are at present a number of projects designed to develop new tests of intelligence which incorporate some of the features outlined in this paper In Geneva, Vinh-Bang and other Piagetian associates have been working on tests of intelligence with a substantial number of tasks (5) Warburton (24) and others at the University of Machester have for some time been working on the construction and validation of the British Intelligence Scale which includes items that test for various Piagetian cognitive stages In a recent study related to the development of the British Intelligence Scale, Lunzer (15) reported the empirical examination of a battery of 21 Piagetian tasks and indicated that with further refinement, a viable developmental cognitive-task battery, which could be used as an intellectual diagnostic device in schools, is highly probable in the foreseeable future Ward and Fitzpatrick (25), in a recent report, elucidated the steps and procedures employed in the construction of the British Intelligence Scale, which is marked by substantial psychological care and zeal on the part of the researchers as evidenced in the series of reformulations on what would be an optimal test of logical reasoning ability When complete, the British Intelligence Scale will have sub-tests designed to measure each of the six Primary Mental Abilities (23) using a substantial array of Piagetian reasoning items The testing time is comparable to that required for the Wechsler Intelligence Scale for Children and the age range for testees is from five to twelve years of age The scale might have been of greater interpretative value had it been exclusively a criterion-referenced ordinal scale of intelligence However, as it stands, it will provide British educators not only with an important research tool, but of even greater importance, it will allow a more informative diagnostic and prognostic intellectual assessment than either the Wechsler Intelligence Scale for Children or the Stanford-Binet Scale

Important work on the construction of an ordinal scale of intelligence has also been carried out at the University of Montreal by Pinard and Laurendeau (18, 19) The context of the scale consists partly of physical tasks dealing with concepts of physics and chemistry Work on the scale has been postponed for an undetermined time (18) as the Montreal researchers scrutinize and confirm basic data collected for each task used to test specific concepts involved in intelligence It is expected that the scale will not be complete for quite a while (17) However, parts of the scale that were discussed by Pinard and Laurendeau (19) have been, and probably will continue to be, studied in research such as that of Dudek *et al* (4) When the Montreal scale and the British Intelligence Scale have been completed, work on the construction of structurally-parallel intelligence tests could then commence, incorporating changes in the types of materials used, in the content areas employed, and in modes of measurement

Other attempts at ordinal intelligence scale construction should be made if for no other reason than that the Montreal scale requires in excess of six hours of individual-testing time, obviously a more administratively-feasible test would be needed for mass use in schools It is entirely possible that ordinal scales of intelligence could be constructed with as few as ten carefully selected items which could be administered in less than thirty minutes

The ramification of such an approach to the measurement of intelligence would be extensive as it would necessitate vast re-experimentation with such topics as the role of genetic factors in the determination of intelligence At present, Jensen (12), among other educational psychologists, has argued that only genetic factors could account for the sizeable differences between the races on tests of intelligence However, genetic components would be likely to account for much less variation among tests scores if custom-made structurally-parallel tests of intelligence were used Also, it seems that, at present, the intelligence of many minority group students is underestimated, since the students are required to manifest intelligent behaviour in contexts that are unfamiliar to them When we have more accurate measures of the optimal level of intellectual functioning of various groups, the differences between groups in intelligence cited by Jensen and others will probably be greatly reduced The issue of the hereditary bases of intelligence is but one educational sphere that is likely to be greatly affected by the approach to the measurement of intelligence discussed in this article

REFERENCES

1 BURT, C The evidence for the concept of intelligence *British Journal of Educational Psychology*, 1955, *25*, 158-177

2 BURT, C Intelligence and heredity Some common misconceptions *The Irish Journal of Education*, 1969, *3*, 75-94

3 DONALDSON, M *A study of children's thinking* London Tavistock, 1963

4 DUDEK, S Z, LESTER, E P, GOLDBERG, J S, and DYER, G B Relationship of Piaget measures to standard intelligence and motor scales *Perceptual and Motor Skills*, 1969, *28*, 351-362

5 FLAVELL, J H *The developmental psychology of Jean Piaget* Princeton, New Jersey Van Nostrand, 1963

6 Flavell, J H, and Wohlwill, J F Formal and functional aspects of cognitive development In Elkind, D, and Flavell, J (Eds), *Studies in cognitive development* New York Oxford University Press, 1969 Pp 67-120

7 Glaser, R Instructional technology and the measurement of learning outcomes Some questions *American Psychologist*, 1963, *18*, 519-521

8 Guilford, J P *The nature of human intelligence* New York McGraw-Hill, 1967

9 Helmstadter, G C *Principles of psychological measurement* New York Appleton-Century-Crofts, 1964

10 Hunt, J McV *Intelligence and experience* New York Ronald Press, 1961

11 Inhelder, B, and Piaget, J *The growth of logical thinking from childhood to adolescence* New York Basic Books, 1958

12 Jensen, A How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, *39*, 1-123

13 Jöreskog, K G Some contributions to maximum likelihood factor analysis *Psychometrika*, 1967, *32*, 443-482

14 Kellaghan, T Abstraction and categorization in African children *International Journal of Psychology*, 1968, *3*, 115-120

15 Lunzer, E A. Construction of a standardized battery of Piagetian tests to assess the development of effective intelligence *Research in Education*, 1970, *3*, 53-72

16 Piaget, J *Psychology of intelligence* Paterson, New Jersey Littlefield, Adams, 1963

17 Pinard, A Personal communication, 1971

18 Pinard, A, and Laurendeau, M *Causal thinking in children* New York International Universities Press, 1962

19 Pinard, A, and Laurendeau, M A scale of mental development based on the theory of Piaget Description of a project *Journal of Research in Science Teaching*, 1964, *2*, 253-260

20 Price-Williams, D R A study concerning concepts of conservation of quantities among primitive children *Acta Psychologica*, 1961, *18*, 297-305

21 Spearman, C *The abilities of man* New York Macmillan, 1927

22 Tanner, J M, and Inhelder, B (Eds) *Discussions on child development Vol I* New York International Universities Press, 1953

23 Thurstone, L L *Primary mental abilities* Chicago University of Chicago Press, 1938

24 Warburton, F W The construction of the new British Intelligence Scale *Bulletin of the British Psychological Society*, 1966, *19*, 59

25 Ward, J, and Fitzpatrick, T F The new British Intelligence Scale Construction of logical items *Research in Education*, 1970, *4*, 1-23

26 Wechsler, D *The measurement of adult intelligence* Baltimore Williams and Wilkins, 1944