Annual review

Intelligence for education: As described by Piaget and measured by psychometrics

Michael Shayer*
King's College, University of London, Cambridge, UK

Two separate paths to the concept of intelligence are discussed: the psychometric path being concerned with the measurement of intelligence, involving the methodology of norm-referenced testing; the path followed by Piaget, and others, addresses from the start the related question of how intelligence can be described, and employs a criterion-referenced methodology. The achievements of psychometrics are briefly described, with an argument that they now remain important tools of what Kuhn called ‘normal science’. The criterion-referenced approach of Piaget and others is described, with evidence from intervention studies that the Genevan descriptions of children-in-action have allowed the choice of contexts within which children can profitably be challenged to go further in their thinking. Hence, Genevan psychology is also now a part of the normal science with important uses, shown both in neo-Piagetian studies and further research stemming from Geneva. Discussion of the ‘Flynn effect’ sheds light on both paths, with problems still unresolved. The argument is then developed that the relevance of neuroscience needs to be discussed to try to decide in what ways it may provide useful insights into intelligence.

The controlling aim of this survey is specifically that account of intelligence that can serve to inform the general education of children and adolescents: the field of concern of the author for, now, 30 years. The paradox exists that the path taken by psychometricians ended in the methodology of Mental Abilities, and the path taken by Geneva ended in the study of children’s notions of Truth and Necessity, with ‘intelligence’ almost forgotten. Historically, neither of the followers of the two paths has been able to communicate with each other, and reasons will be suggested for this. More importantly, the case will be explored that only by regarding the two paths as complementarities is it possible to recover an understanding of intelligence that will serve teachers in schools.

*Correspondence should be addressed to Michael Shayer, 16 Fen End, Over, Cambridge CB24 5NE, UK (e-mail: m.shayer@ukonline.co.uk).
**Criterion and norm-referenced testing**

**Criterion-referencing**

This is most conveniently illustrated by the most-used criterion-referenced tests in England and Wales: those used to assess the achievements of 14-year-olds in school in science and mathematics (Key Stage 3 Standard Assessment Tests or SATs). The National Curriculum (NC) – originally described in 1988 in 10 levels – is currently described in 8 levels ranging from what might be found from Year 1 in Primary school to Year 11 in Secondary school. At each level there are enough Statements of Attainment (SoAs) to specify, for example, competence in Mathematics with regard to Number; Shape, Space, and Measure; Algebra; Handling Data; and Using and Applying Mathematics, each being assumed to have equivalent degrees of difficulty.

The mandatory Key Stage 3 Maths SATs for all 14-year-olds therefore constitute some 220 question items each of which is tied specifically to one or more of the SoAs, ranging across NC levels 3 to 8. Each pupil is assessed as being at the highest NC level at which they have achieved a 67% success, items achieved at higher levels being added in.

Thus the Key Stage tests, if the questions are validly written to express the SoAs, are referenced to the criteria of the National Curriculum. If ‘intelligence’ could be defined in an equivalent richness of description then it would be possible to devise a criterion-referenced test for it.

**Norm-referencing**

This constitutes the approach of all psychometric testing since the early 1920s. Whether it be used to investigate and measure all the mental abilities that are related to intelligence, or whether it is applied to performance scales such as the Wechsler Arithmetic, the essence of the method is to arrive at a measure of a child’s proficiency by relating his test score to a sample representing the whole population – the normative group – after allowing for his/her age. The measurement is that of the extent to which the child is above or below or at the mean of the normative group of the same age. This then permits quantitative correlational methods to be used in investigating intelligence.

Although this distinction is the main organizing principle of this paper, first the contribution of Binet needs a separate account, since his work spans both approaches.

**The contribution of Binet to the concept of intelligence**

Binet (1905; chapter II in Binet & Simon, 1980, p. 37) stated:

> Our purpose is to be able to measure the intellectual capacity of a child who is brought to us in order to know whether he is normal or retarded.

He goes on to say (pp. 39, 40):

> Our purpose is in no way to study, analyse, or set forth the aptitudes of those of inferior intelligence . . . Here we shall limit ourselves to the measuring of their general intelligence.

The fundamental idea of this method is the establishment of what we shall call a measuring scale of intelligence. This scale is composed of a series of tests of increasing difficulty, starting from the lowest intellectual level that can be observed, and ending with that of average normal intelligence. Each group in the scale corresponds to a different mental level.

---

1 NB it is important to distinguish between norms of performance, as referred to here, and norms of thinking, discussed later in the context of Piaget’s work.
Like Piaget later he claims (p. 42):

We believe that we have succeeded in completely disregarding the acquired information of the subject. We give him nothing to read, nothing to write, and submit him to no test in which he might succeed by means of rote learning. . . . It is simply the level of his natural intelligence that is taken into account.

In the 1905 version of his test (Binet & Simon, 1980, chapter II) he starts out with 30 subtests going a fair way round the spectrum of mental abilities with the intention of assessing the normal development of children in the age range of 3–12 years. At this point, there is only description of work-in-progress (Binet & Simon, 1980, chapter III) but it is clear that the intention is to arrive at selection of items, success on which will place a child at the level of the normal 3-, 4-, or 5-year-old and so on. ‘Normal’ for a child here is defined as within ± 1 mental age year of what is normal for the child’s age. Hence, for diagnostic purposes, it would be possible to make statements like ‘this 8-year-old functions at the level of the normal 5-year-old, has a mental age of 5’. This is not measurement in the psychometric sense: each of the year-levels is described in terms specifically of what the child can do. It is much closer to criterion-referenced testing, and Binet resisted all attempts to put numbers on his test results.

Binet and Simon (1908; Binet & Simon, 1980, chapter IV) report the development and revision of the 1905 test. But here it should be noted that, when Binet wished to begin study of intellectual development he first, like Piaget 20 years later, made a detailed study of the thinking of his own two young daughters. In this paper (p. 183) he remarks:

The child differs from the adult not only in the degree and quantity of his intelligence, but also in its form. What this childish form of intelligence is, we do not yet know. In our actual experiments we have only caught glimpses of it.

Thus it is easier to see how Piaget found the work of Binet and Simon a good starting-point for his own work. Indeed, the detail on levels of children’s definitions in Binet and Simon (1980, pp. 204, 205) might have been taken as part of chapter 6, pp. 141–145 on Analogies, in Piaget (2001), where Piaget does specifically acknowledge Binet. On the basis of three more years of experience, it was then possible for Binet and Simon to specify either five or six specific tests – each given with lists of satisfactory and unsatisfactory responses close to the detail of a Piagetian protocol – for each mental year of age (Binet & Simon, 1980, Table, pp. 238, 239). A success criterion is included for all bar one of the tests, but with this further proviso (p. 244):

. . . when once the intellectual level of a child is fixed, give him the benefit of an advance of one year every time he passes at least five of the tests beyond his level, and the benefit of an advance of two years if he has passed at least ten above his level.

Thus, he was already aware of differential mental abilities: indeed in this remarkable paper he touches on every issue relating to ‘intelligence’. For example, on p. 262 under ‘Uses’, this use of the method:

. . . it demands that one establish by careful investigations the normal evolution of a child’s intelligence in order to make all our programs and methods of instruction conform to that evolution, once it is known.

In the last 1911 paper (Binet & Simon, 1980, chapter V), some detail of the revision of test items is given, but the bulk of the paper is given over to discussing the use of various
forms of assessment by teachers in schools. He does (p. 301) make a statement very close to that of Piaget's (1918, pp. 150, 173, 176–178) initial concept of equilibration:

our favourite theory; the intelligence\(^2\) marks itself by the best possible adaptation of the individual to his environment.

The psychometric path

However, in 1917 it was necessary quickly to assess the mental capacity of new US Army recruits, so Yerkes (Yoakum & Yerkes, 1920) developed, from the Stanford–Binet form (Terman et al., 1915), the first test that could be answered in writing by a large group: the Army Alpha test – if they could not read there was a Beta test administered individually. Here the total score put the recruit into five levels from A to E. Only those graded C and above were considered as possible officers, and E and below were rejected for service.

Terman et al. (1915) also developed the concept of the IQ measure. Given norms on children of each year of age, with their mean defined as ‘Mental Age’, then the IQ of a specific child was defined as:

\[
\text{IQ} = \frac{\text{Mental age}}{\text{Chronological age}} \times 100
\]

An IQ of 100 would therefore mean the child was exactly at the national average.

The Terman definition gives rise to various measurement anomalies, as well as not being applicable to adults, so the definition currently used (Rust & Golombok, 1999, p. 84) is:

\[
\text{IQ} = 100 + \frac{15 \times (\text{Raw score} - \text{population mean score})}{\text{Population SD}}
\]

with the Population mean estimated from a representative sample of children of the same chronological age. This is the same as that introduced in 1939 for the Wechsler tests.

The achievement of psychometrics: 1 – Measurement

Psychometrics has featured the greatest investment in ‘numbers’ among attempts to make psychology scientific. Indeed, two different aspect of numbers have been used: on the one hand, an investment in statistics in the development of test theory and on the other the development of correlational methods in the art of factor analysis used in giving meaning to data derived from the better tests delivered by test theory. Yet, this hides the great paradox hidden below this approach to understanding intelligence: the result of this application of quantitative methods is to narrow the focus of attention only on to the qualitative aspects of all the different kinds of thinking and actions that can be described in terms of test items. This paradox will now be explored.

One needs to remember that both in the USA and the UK the dominant model in psychology from 1920 to the late 1960s was behaviourism, with the implicit assumption that for a scientific approach only the response to a stimulus may be observed (S-R): any model of an intermediate process was described as mere mentalism, and was implicitly prohibited.

\(^2\)‘intelligence’ – translation as given in Binet and Simon (1980).
In test construction, each item is a *stimulus* with the correct *response* defined. A person’s competence on the test is assessed in terms of how many correct responses they make. But setting out on the road to measurement in this way is roughly analogous to that of a skier heading down a steep piste in the snow: you think you are in control but very soon you forget everything except how to manage the next bend or drop coming up – the piste controls you. You may start with an idea of what it is you set out to test – in the case of Spearman (1927) and intelligence, it was ‘the eduction of relations and the eduction of correlates’ – and then use that to write some possible test items. Remember that you are forbidden to try to describe mental processes. When you give the test to a sample of children, and look at the results from the point of view of test theory you realize that some of the items do not work. It may be because they are either too difficult or too easy, or it may be that they do not correlate well with the rest of the items – that is some children may succeed on them while failing on most of the other items, and others may fail on them while succeeding on most of the other items. So you either throw the items out, or you try to save them by rewriting them and trying them on a new sample (this rarely works: more economical of effort is to write some new items). Already the mathematical apparatus is controlling you. But then comes the question of how to convert the number of items a child has got right to some kind of meaningful measurement – some scale that has measurement properties like that of a ruler.

At this point, the end of this complex process of investigating intelligence is a simple score, say 110. For the purpose of a teacher in school, how much information has been imparted? Even if it is a performance test focused on mathematics rather than a more general test of intelligence, all you know is that the child is at the 75th percentile for his age, and for many teachers even that information is too number-rich for them to process. You haven’t been told what mathematics he can do, and there is no way you can estimate from his present measurement at 9 years of age what mathematics he is likely to be able to cope with at 11.

**The achievement of psychometrics: 2 – Mental abilities and intelligence**

Here the metaphor of the piste is taken up again. What is it that the psychometric path can achieve, and what is it that it cannot? A comparison might be drawn between the situation of the physicists in the 1920s and the two separate paths taken in psychology at the same time. Just as the practitioners of the new wave-mechanical model of atomic systems adopted the view that only the actual measurements – taken together with a probabilistic mathematical model of their connection – were to be mentioned, so too the behaviourist approach to psychology only allowed the response to the stimulus (S-R), plus any mathematical models of their connection to be mentioned. Hence the complementarity principle of dealing with particles and waves might be paralleled by the norm-referenced and criterion-referenced modelling of intelligence in psychology.

From the very beginning, with the Binet–Simon test, it was assumed that many different aspects of behaviour needed to be measured in order to estimate intelligence. Given data on children of different ages any attempt to be precise on what these aspects have in common leads to almost zero communication, as Spearman’s (1927) late summary of the art witnesses. Cattell (1971, p. 26) tells the story of Spearman rattling off in a lecture strings of mathematical formulae, and giving him a piece of chalk in the hope that he would record them on the board. Spearman kept the piece of chalk in his hand, and concluded the lecture by saying ‘And this is what I call the theory of “g” and writing
the one expression ‘g’ on the board. The questions that can be asked as the piste is embarked upon are, what is the relation between these aspects and what model of these relations can be created? The research programme thus defined lasted 77 years (Terman et al., 1915, to Carroll, 1993). Statistical methods and the development of test theory increased the precision of the measurements, and the development of correlational methods enabled the strength of relations between the qualitative definitions of the different aspects of intelligent behaviour to be described and then modelled. Cattell (1971, p. 10) in speaking of defining intelligence said ‘...one must first acquire a deep suspicion of words’ and ‘...the approach of making up subjective, armchair definitions of intelligence is foredoomed, logically and methodologically’. Carroll (1993, p. 36) devoted just 1 page out of 714 to summarizing others’ words before throwing away the possibility of that path as being useful.

Such a research programme follows from its implicit assumptions. Given the Stimulus–Organism–Response model of experimental psychology in the laboratory, Spearman and Thorndike would naturally proceed as was later described by Cattell (1971, pp. 9, 53):

The scientist... has to proceed empirically and iteratively. That is to say, he starts with a rough definition of what he is looking for and gradually reshapes it as he begins to see what is really there.

First, a wide variety of tests of different behaviours are constructed, given to a good sample of children and the data are then subjected to correlational analysis.

The faculty psychologist of a hundred years ago found a word in the dictionary (or floating in his mind) and then described at length the manifestation of a unitary mental capacity which conformed to it. The factor analyst, or correlation analyst, does something very different when he finds the unitary pattern first, and then describes and names it according to its form.

That is, there is an experimental method. For each definable ability, items of varied supposed difficulty and of different expressions of that ability are written. Test theory then enables the test to be purged of unsatisfactory items, and enough items are written to allow the internal consistency to be high enough, 0.8+ perhaps. The problem then arises, what can you do when a sufficient variety of such tests have been created (Cattell, 1971, p. 29, lists 22 Primary Abilities, ranging from Thurstone’s Verbal Ability and Numerical Ability through Cattell’s Ideational Fluency to Guilford’s Motor Speed)? The answer is that you look, by correlational analysis, to see what light they show on the original enterprise of investigating intelligence in a scientific way. ‘Intelligence’ then disappears (it is not listed in Cattell’s Glossary of technical terms, and had already disappeared long since into Spearman’s ‘g’) and the complement, quantitative relationships between the Mental Abilities established by factor-analytic methods, then appears as the end-product of the research programme. Suitable combinations of these tests of mental abilities can then find practical use, whether it be the Stanford–Binet, the various Wechslers, the Thurstone PMA and so forth, as batteries that laymen will take as tests of ‘general intelligence’.

Cattell (1971) shows in great detail the impressive achievement of the psychometric path. His Factor-analytical programme of research yielded two general g factors, fluid intelligence and crystallized intelligence, abstracted from all the individual tests of different mental abilities. His Culture-Fair test battery loaded mainly on gf. Quoting a general correlation of g measurements with tests of school learning between 0.5 and 0.7, and arguing that fluid intelligence was also related to school
learning measured by tests of crystallized intelligence, he did once (but only once, pp. 440, 441) float the idea that there might be such a thing as constraints imposed by absolute levels of \( g_f \):

... whatever the amount of teaching, the individual could not acquire real judgement response to the right abstraction of situational elements, beyond that attainable from his particular level of \( g_f \).

But he also floated the Vygotskian notion (p. 443) that achievements in learning, even if initially learned without underlying insight, can serve as a springboard for 'creative steps taken by virtue of \( g_f \).

Cattell, and Catell only to my knowledge among psychometricians, recurred several times through the book (pp. 140, 145, 176, 456) to the enormously wide spread of measured mental abilities in the child and adult population, with the standard deviation of fluid intelligence tests being 50% greater than that of crystallized intelligence – a spread that later was published on Piagetian tests by the CSMS programme in the mid-1970s (see below).

He also (pp. 446–467) – before Flynn’s work – commenting on the rise in the crystallized general intelligence over the last 50 years, suggested the desirability of having some test that could serve as an ‘absolute’ standard for comparison over generations. Yet (p. 457), he thought this rise may well have been ‘... accompanied by a loss of intellectual skills in many other directions’, a thought echoed in detail in Flynn (2007).

Finally, whereas fluid intelligence appears to rise to a maximum by 16 (p. 144) and starts to decline from early in the 20s, crystallized intelligence (p. 169) rises sharply until about 18, but thereafter continues to rise gradually until some time between 50 and 60.

The view taken here is that on ‘normal science’ Kuhn is right and Lakatos is wrong. At any one time point most of the science taught and practiced most of the time by most scientists is what Kuhn describes as normal science. But Lakatos is right in pointing out how slowly new research ideas permeate into the science population, and Kuhn’s idea of ‘revolutions’ hardly describes what actually happens. Max Planck said you just have to wait until the older professors die.

By the time of Cattell’s (1971) summary the research programme of psychometrics had virtually exhausted its heuristic power, although Carroll’s (1993) truly monumental re-analysis of some 460 datasets could be said to have sculpted the final tombstone. His final three-stratum model appears to contain all the information that psychometrics has delivered, with a general factor at the top and in addition to \( g_f \) and \( g_c \) on the next level, there are others such as a General Memory and a Visual Perception factor.

It is no detriment to this research programme to say that it is over: on the contrary, it has entered permanently into the normal science of psychology, with its current usages in the fields of neuropsychology, clinical psychology, and educational psychology.

**The criterion-referenced path**

**The school of Piaget**

Piaget begun his intellectual life in his teens in biology, accompanied by a strong interest in philosophy (Piaget, 1952). At the age of 21 he was given an appointment by Simon in Paris, in part involving norming the new Burt test for use in France. He quickly tired of this, becoming more interested in the reasons that children gave both for their mistakes and their successes.
Instead of starting with test items as stimulus, with response correct or incorrect, Piaget almost at once diverted from the quantitative path initiated by Binet. He asked the question, how, qualitatively, and then hierarchically, can we describe intelligent activity in child and adult? More importantly, he also asked the question, how can we describe the development of intelligent activity in the human, from the time of birth onward? But these questions cannot be addressed without a description of knowledge in itself, and for this it is necessary to turn to some 2,400 years of work in philosophy from Aristotle, through in particular Kant; and inasmuch as logic, governing processes of inference, is an essential part of the description, to Frege onwards. Hence Piaget’s description of his life’s work as genetic epistemology, better named developmental epistemology (Smith, 2006, p. 115) how humans come to be able to act in the world and come to have knowledge.

Piaget’s clarification of his research programme also needs describing developmentally, even if seeds of it can be perceived early on. In the 20s – Phase 1 – he worked essentially in an empirical way, investigating the language and thought of the child, their judgment and reasoning, and their ideas of causality.

Although he was at this time, as always, interested in the logic of the children, in Phase 1 he made much use of sociological/biological descriptions of the forces acting on the child. He was establishing his experimental method, which is essentially that of the biologist, and none the worse for so being. Whereas for the norm-referenced approach to psychology the sample needs to be large enough to establish a mean and standard deviation representative of the population (1,000–2,000 for each year of age), the biologist needs only enough ‘good specimens’ as will suffice to describe variations in the species. Given 20 well-chosen children (neither geniuses nor those defective in any way) it is highly unlikely that any essential behaviour will be missed. By choosing a range of ages during which a particular kind of thinking is developing, the whole of what Vygotsky called ‘the zone of proximal development’ will get described, from the very first beginnings through to the confident practice of the scheme studied. The fascination of the method is that – unlike much of reported science – Piaget gives the reader the Minute Particulars of what each child did and said so that the reader can make his own abstraction to check the fit of Piaget’s interpretation. Particularly in Phase 1 of Piaget’s work, the individual interview features a dialogue with the child to see how he/she justifies the act they have just executed, be it verbal or gestural. Another achievement in this phase (continued for the rest of his life) is that much of the Genovan method was actually executed by six or more assistants (including his wife-to-be) so that collaboration was built in to the research.

Phase 2 of his work started in 1925 with the study of his own three children from 0 days 2 hours onwards until 1932 (The Origin of Intelligence, Piaget, 1936, 1953). Somewhere around this time, Susan Isaacs – possibly on Piaget’s visit to Cambridge in 1928, and later published in Mind as Isaacs (1930a) – had criticized the Language and Thought of the Child (Piaget, 1925) as having been overmuch concerned with verbal transactions, and also being too pessimistic a description of children’s development. But at the Malting House school in Cambridge, the mean IQ (on the Mental Age definition) of the children was 131 (Isaacs, 1930b, p. 96) so it is hardly surprising that they showed behaviours two years younger than Piaget’s children at the Maison de Petits in Geneva. With his own babies he was forced to record just their actions, his own interactions with

---

3 ‘The General Good is the plea of the Scoundrel and the Hypocrite: Good is done in Minute Particulars’. William Blake.
them designed to elicit their behaviour, and their further reactions. Henceforth his concern would be as much for the implicit logic of action⁴ as of verbalization, and the Genevan method was modified accordingly. This phase led into 20 years of collaborative work with Szeminska and Inhelder, beginning with Number (Piaget & Szeminska, 1952), Space (Piaget & Inhelder, 1976), Geometry (Piaget, Inhelder, & Szeminska, 1941), Movement and Speed (Piaget, 1970), Physical Quantities (Piaget & Inhelder, 1974), Early Growth of Logic (EGL) (Inhelder & Piaget, 1964), and Growth of Logical Thinking (GLT) (Inhelder & Piaget, 1958).

Given that developmental epistemology (DE) is the aim - the third or middle way between psychologist and logicism (see Smith, 2006, pp. 107–116, for the history of this distinction) - it is necessary to ask what relation this aim has to developmental psychology. The claim is that, inasmuch as the empirical study of how children develop norms of thinking necessarily at the same time describes what an applied researcher would want to know about children’s ability to understand the learning they are presented with, the Genevan Phase 2 work provided much of what the normal science of developmental psychology (DP) needs. After the Origin of Intelligence had refined and established the method, it was necessary to go over the ground of the development of thought, firstly again properly with the younger children, and then finally, to research how the process continued in adolescence.

But DE and DP are not identical, even though the area of overlap in Phase 2 is considerable. In his Phase 3 work from 1965, the paths of DE and DP then clearly diverge again, with the area of overlap much less, though never zero. Piaget (2006, pp. 7, 8) reverts to the epistemological issues in his mind from the early 20s:

The key question for analysis is the examination of what subjects regard as proof or ‘reason’ for what they regard as truth.

Phase 2, nevertheless, has such consequences for the contribution of psychology to intelligence that it needs a separate description.

Geneva and developmental psychology

Looked at from the point of view of psychometrics, the key works of this period cover a substantial part of the spectrum of mental abilities. But they were achieved through collaboration with two workers, formidable in their own right, Alina Szeminska and Bärbel Inhelder, whose contribution could be read as keeping Piaget’s feet on the ground. Essentially they were concerned with the specifics, the ‘Minute Particulars’ of children’s behaviour, empirically ascertained and recorded. In the case of GLT (Inhelder & Piaget, 1958) all the experimental work was done independently as Inhelder’s thesis (never later published)⁵ on the development of inductive thinking in children and adolescents. As such, it is an amazing tour-de-force, as nearly every chapter is a well-chosen context of scientific concepts going just high enough, but not too high, to elicit the levels of thinking that the more able adolescents are capable of. Her experimental intuition enabled thinking at equivalent levels of complexity to be described in many different contexts, covering schemes featured both in the physical and the biological sciences, as can be seen in Table 1.

⁴ It is important from the outset to point out that any mention of ‘logic of action’ etc. should not be confused with von Wright’s account of deontic logic. Modal logics were explicitly not used by Piaget (though see later with regard to Piaget & Garcia, 1987).
⁵ It does exist, though, filed in the Archives at Geneva. Les conduits experimentales chez l’enfant et chez l’adolescent. www.unige.ch/piaget/
As part of the work of the CSMS programme, Shayer, Kücheman, and Wylam selected five chapters to investigate: Pendulum, Equilibrium in the Balance, Inclined Plane, Chemical Combinations, and Flexible Rods. Using the behavioural descriptions Inhelder/Piaget had provided at every level from mature concrete (2B) to mature formal (3B) in each chapter, and using standard psychometric test theory and practice, they constructed ‘Demonstrated Class Tasks’ with 14–17 items for each task. The same apparatus used by Inhelder was used, and the school classes tested were taken through all the key aspects of each task featured in Inhelder’s detailed descriptions, using her success criteria at each of the levels. The internal consistency of the tests ranged from 0.76 to 0.86. When tested on a comprehensive school sample of 100 boys, their mean Piagetian level scores were within sampling error on all five tasks (Shayer, 1978, p. 227), indicating that the Inhelder/Piaget scaling of the tasks was equivalent between each. The same did not apply to the sample of girls, where their performance on two of the tasks, Equilibrium in the Balance and Inclined Plane, was substantially lower. This was attributed to a sex differential already found on the Volume & Heaviness task (Shayer, Coe, & Ginsburg, 2007).

Bearing in mind the earlier praise for the Genevan method, and the assertion that a sample of 20 ‘good specimens’ was enough to elicit significant behaviours, this may appear a perverse procedure. But it involves just the same principles as, for example, constructing a criterion-referenced test for the Key Stage 3 SATs for science and mathematics, given the National Curriculum. Just as the SATs place a pupil at one of the NC levels from NC 3 to 8, given the NC hierarchy, so too given the Genevan hierarchy, the Class Tasks can place a child’s performance on a 6-point scale ranging from early concrete (2A) to mature formal (3B). Norm-referenced tests, in principle, lack such a criterion. One of the key members of the original group who constructed the National Curriculum hierarchy of levels and SoAs for Mathematics, Margaret Brown, was also a key member of the CSMS research team.

Showing that the detail exhibited in Number, Physical Quantities, Space, Geometry, Movement, and Speed, EGL and GLT (Inhelder & Piaget, 1958, 1964; Piaget & Inhelder, 1974, 1976; Piaget & Szeminska, 1941, 1952; Piaget et al., 1941), was rich enough to describe every level of performance in the contexts of science from early concrete through to mature formal, Shayer (1972) was able to assess the level of thinking implicit in each of the lessons of the Nuffield O-level courses in Chemistry, Physics, and Biology on two criteria: the minimum Piagetian level a pupil need be for their interest to be engaged and the minimum level a pupil need be to comprehend the concepts at that point in the course. From this analysis, he also showed that the Nuffield courses had achieved better understanding of the underlying science at the expense of increasing the intellectual demand level of the lessons. The taxonomies used in this work were published in Shayer

---

Table 1. Schemata studied in the growth of logical thinking

<table>
<thead>
<tr>
<th>Biological and social science</th>
<th>Physical sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control of variables</td>
<td>Coordination of frames of reference</td>
</tr>
<tr>
<td>Exclusion of irrelevant variables</td>
<td>Multiplicative compensation</td>
</tr>
<tr>
<td>Probability</td>
<td>Equilibrium of physical systems</td>
</tr>
<tr>
<td>Correlation</td>
<td>Proportional thinking</td>
</tr>
<tr>
<td>Combinatorial thinking</td>
<td>Physical conservations involving ‘models’</td>
</tr>
</tbody>
</table>

As part of the work of the CSMS programme, Shayer, Kücheman, and Wylam selected five chapters to investigate: Pendulum, Equilibrium in the Balance, Inclined Plane, Chemical Combinations, and Flexible Rods. Using the behavioural descriptions Inhelder/Piaget had provided at every level from mature concrete (2B) to mature formal (3B) in each chapter, and using standard psychometric test theory and practice, they constructed ‘Demonstrated Class Tasks’ with 14–17 items for each task. The same apparatus used by Inhelder was used, and the school classes tested were taken through all the key aspects of each task featured in Inhelder’s detailed descriptions, using her success criteria at each of the levels. The internal consistency of the tests ranged from 0.76 to 0.86. When tested on a comprehensive school sample of 100 boys, their mean Piagetian level scores were within sampling error on all five tasks (Shayer, 1978, p. 227), indicating that the Inhelder/Piaget scaling of the tasks was equivalent between each. The same did not apply to the sample of girls, where their performance on two of the tasks, Equilibrium in the Balance and Inclined Plane, was substantially lower. This was attributed to a sex differential already found on the Volume & Heaviness task (Shayer, Coe, & Ginsburg, 2007).

Bearing in mind the earlier praise for the Genevan method, and the assertion that a sample of 20 ‘good specimens’ was enough to elicit significant behaviours, this may appear a perverse procedure. But it involves just the same principles as, for example, constructing a criterion-referenced test for the Key Stage 3 SATs for science and mathematics, given the National Curriculum. Just as the SATs place a pupil at one of the NC levels from NC 3 to 8, given the NC hierarchy, so too given the Genevan hierarchy, the Class Tasks can place a child’s performance on a 6-point scale ranging from early concrete (2A) to mature formal (3B). Norm-referenced tests, in principle, lack such a criterion. One of the key members of the original group who constructed the National Curriculum hierarchy of levels and SoAs for Mathematics, Margaret Brown, was also a key member of the CSMS research team.

Showing that the detail exhibited in Number, Physical Quantities, Space, Geometry, Movement, and Speed, EGL and GLT (Inhelder & Piaget, 1958, 1964; Piaget & Inhelder, 1974, 1976; Piaget & Szeminska, 1941, 1952; Piaget et al., 1941), was rich enough to describe every level of performance in the contexts of science from early concrete through to mature formal, Shayer (1972) was able to assess the level of thinking implicit in each of the lessons of the Nuffield O-level courses in Chemistry, Physics, and Biology on two criteria: the minimum Piagetian level a pupil need be for their interest to be engaged and the minimum level a pupil need be to comprehend the concepts at that point in the course. From this analysis, he also showed that the Nuffield courses had achieved better understanding of the underlying science at the expense of increasing the intellectual demand level of the lessons. The taxonomies used in this work were published in Shayer

---

and Adey (1981). Later, others used the same approach to assess the levels of cognitive demand in English Literature (Fusco, 1983), Music (Evans, Petrie, & McLoughlin, 2006), Dramatic Arts (Alty & Pout, 2006), and Visual Arts (Leighton & Quinn, 2006).

Given the justification of deriving criterion-referenced tests, based on the well-researched behavioural criteria reported from Geneva, then it was possible, during the CSMS research programme, to investigate the population as a whole. Remember that for the primary schoolwork Piaget used classes representative of the population as a whole – hence some two years behind on average what Susan Isaacs found in the Malting House school – whereas the Inhelder work used older subjects from the gymnasia, a highly selected sample. That left open the question whether the upper levels of thinking described in Inhelder’s ‘good specimens’ were actually achieved by all by the end of adolescence. A major part of the CSMS programme was to investigate just this.

Three Piagetian Class Tasks (NFER, 1979) were constructed and validated using standard psychometric test practice: Spatial Relations, based on Space, to pick up every level from 1B, intuitive, through 2A, early concrete, to 2B*, concrete generalization; Volume & Heaviness, based on Quantities and Number, covering early concrete to early formal; and Pendulum, based on GLT chapter 4, to assess the higher levels from mature concrete to mature formal. During one school year, a representative sample of 10,000 pupils between the age of 10 and 16 years, the younger ones from middle schools, were assessed on the three tests, giving a sample ranging from 1,000 to 2,000 for each year of age. Figure 1 shows the results of the CSMS survey as a whole. Figure 2 shows the distribution for the 14-year-olds.

These show an essential feature of the population statistics that Cattell (1971, p. 456) noticed (speaking of the adult population): ‘. . . the range of mental age still goes from three or four years to the equivalent of an abstract mental age of twenty-five years’. The spread here is enormously wider than that on any other kind of developmental data on children. Lest there be any question of there being something strange about Piagetian test results, Figure 3 shows data on 14-year-olds for Key Stage 3 Maths for 2000 showing just the same range as is shown in Figure 2. The other feature of the CSMS survey was that only 14% of the 16-year-olds exhibited mature formal operations, with only 31% at the early formal level. Yet, a scan of the Social Science Citation index from the early 1980s onwards records not a single reference to the published CSMS data (Shayer, Kuchemann, & Wylam, 1976; Shayer & Wylam, 1978) anywhere in any book or article – other than self-citations or citation by colleagues. These facts have not entered the consciousness of the world of psychology.

Here the technology of norm-referencing in the psychometric sense is used to discover the population distribution of the norms defined through Genevan research in the criterion-referencing sense. If Carroll (1993) marks the end of the psychometric research programme of representing Mental Abilities as a contribution to ‘intelligence’, then the Phase 2 of the Genevan work marks the end of the sub-research programme of Piaget’s work as a contribution to developmental psychology. Both should be regarded as complementary contributions to the investigation of intelligence: from Geneva only the nebulous concept of ‘décâlage’ stands in for any information about the relation between mental abilities that the psychometric path supplies in abundance.

**Piaget’s research programme**

Lakatos (1978, pp. 47–52) distinguishes between the negative heuristic that is the unanalysable ‘hard core’ of research programmes, and the positive heuristic that
provides a progression of models derived from but protecting the hard core. Just as for psychometrics the ‘hard core’ S-R then leads to the positive heuristics of test theory and factor analysis, so with Piaget the ‘hard core’ of *equilibration* relates to both his experimental method and to his various attempts at explanatory models.

Starting strictly from Hume’s principle of ‘You can’t get an “ought” from an “is”’, Smith (2006, p. 113–116) gives a detailed discussion of the paradoxes underlying Piaget’s practice of developmental epistemology. Given that:

‘Logic is the study of true reasoning [and so the question arises] how the child controls the *truth* of these deductions, how the idea of truth is successfully gained in the first place’ (Piaget, 1923, p. 57, cited in Smith, 2003, p. 43)

and there are:

‘…facts in experience permitting the observation that a particular agent considers him- or herself to be obligated by a norm, irrespective of its validity from the observer’s point of view’ (Piaget, 1950, p. 30), DE then requires:

‘…the coordination of factual data and normative validities [placed] in correspondence with each other without reducing one to the other.’ (Piaget, 1966, p. 153)
Piaget’s ‘middle way’ is that a scientific and empirical study is perfectly possible of the emergence and development of the norms of thinking – whether these be the Phase 2 schemes of concrete and formal operations familiar to developmental psychologists, or if they are the norms of truth and necessity studied in Phase 3. This of course means that he needed to draw both from epistemology and psychology, since only from insights derived from the first would he be able to know what to look for in the latter. This parallels his view of children in that:

‘... [a human agent] is always “norm-laden”’ (Piaget, 1965, p. 159)

and that:

‘... even amongst our youngest subjects, a physical fact is recorded only within a logico-mathematical framework, however elementary it may be’ (Piaget, 2001, p. 320)

Thus, if the norms of thinking are not derived directly from empirical observations by children how can scientific investigation proceed? Piaget’s answer is that of a biologist: an empirical study of their ontogenesis in children will reveal much that we need to know, and by taking successive ‘snapshots’ of the development of norms it will be possible to see how the interaction of children with their world becomes more and more powerful. But there is more to his programme than that. If, for example, he describes in one of his own babies her construction of an object as deriving from a coinciding of seeing with her gaze, and her hand and arm waving at random meeting the object, then we have a notion of logic very different from that of (most) philosophers.
From the outset his account of ‘knowing’ is a description of the activity of children within their environment, whether it be that of assimilation or of accommodation.

In a typical Genevan investigation, such as on the conservation of quantities with 5 to 8-year-olds, children are first presented with a task in which – as in everyday life – they first get a perceptual or causal view of the situation, and then further questions and suggestions with respect to children’s action reveal the extent to which they have progressed towards confident use of the norm of conservation. From the well-chosen small sample, every stage of progression towards confident use of the norm can be described – sometimes, indeed, the progression can be seen occurring in the individual child during the investigation. Hence the notion of ‘cognitive conflict’ as a driver of cognitive development: the initial reading of the situation for the child prompts the question, why? and in some provides the emotional stimulus for further action by the child towards a solution.

Norms, so conceived, are what enables children, with more and more power, to act within their environment. In the concrete operational period – ‘operations’ because some of the action is now internalized – the norms are those of classification, seriation, the cardinal and ordinal properties of numbers, conservations, spatial relations, and so on. These enrich children’s descriptive powers as they act within the world and give rise to simple notions of causation. At the formal operations stage the norms are those 8 schemata of hypothetico-deductive thought listed in Table 1.

Such a view raises two issues, to be dealt with successively. In the Lakatosian sense, is there independent evidence to strengthen the positive heuristic of Piaget’s programme? And, was Piaget’s symbolic logic account of formal operations a good move in his research programme? These questions are addressed in the ensuing sections.
Piaget’s lifelong use of the notion of equilibration is a typical biologist’s idea relating to the 19th century paradox that homeostatic life-processes appear to contradict the second law of thermodynamics (there is of course an answer to the paradox, that living organisms only do this at the expense of creating even more disorder elsewhere). Such equilibria are always dynamic interactions with the surrounding environment. Piaget’s contribution to psychology is to extend this concept through from the interactions of newly born babies with their immediate environment to the actions of thought – including verbal activity - first of children and then to the more sophisticated reasoning of gifted adolescents. It is the logic of action with the environment that he describes, right up to the point where the action is entirely internalized, as in reflective abstraction on previous actions.

This extends implicitly to the experimental method, the main positive heuristic of his scientific enterprise. Such a notion of equilibration implies the contexts that it is embedded in, which implies choice of a variety of relevant and exemplifying contexts. This in turn implies neither a process of generalization nor inductive thinking, but the unusual kind of abstraction by which mathematicians relate the theory of groups to ‘realizations’ of groups. For instance, the Klein 4-group models what the system of elements of addition of numbers (requiring the invention of the number 0) have in common with the system of the three-dimensional half-turns of a cigar box, despite the fact that there is no causal relation between them. Piaget’s genius lies in his imaginative ability to create experimental contexts for exploring every aspect of a particular kind of cognitive action that his abstract sense of equilibration suggests may be fruitful.7

Another facet of this is that the Genevan workers shared an intuitive sense – never reducible to a formula – of equivalent levels of logical complexity across many different contexts that has rarely been found to be inaccurate in replication studies.

It is just because Piaget could never satisfactorily define ‘equilibration’ that from beginning to end it served him as a ‘hard core’, perpetually seeking new contexts to aid the expression of its multitude of faces – pace Bruner (1959, p. 365) who wrote:

I can see nothing served either in the design of experiments or in the thrust of the general theory that is contributed by the idea of equilibrium.

On the other hand he did acknowledge (p. 363) that:

Piaget’s genius... consists in designing intellectual tasks that externalise thought into action in such a way that one is able to infer the logical assumptions on which the action was based...

In effect, what was labelled ‘g’ in the psychometric literature was the lifetime study of Piaget’s programme.

Evidence from intervention studies
The CASE intervention (1983–1987)8 was a two-year project for 12- to 14-year-olds based on the norms of formal operations. Familiar science contexts were found that embedded the schemata. Thus, for Control and Exclusion of Variables pupils were provided with hollow tubes of both glass and plastic, of differing widths and lengths,

7 Dienes (1963) employed a similar genius in inventing ‘realizations’ of groups.
and asked to find out what governed the pitch of the note obtained by blowing across the end. The classroom atmosphere envisaged was one where the teacher mediates the process of reflective abstraction and conflict resolution. The change is attempted through guided reflective abstraction in line with Piaget, in a context of Vygotskian-derived collaborative learning (Shayer, 2003). Although the CASE lessons were taught only at the rate of one every 2 weeks over a two-year period, the teachers were encouraged to regard them as a sample scheme from which they could modify the rest of their science teaching, and this ‘bridging’ process is probably the reason for their substantial effect. Piagetian tests, which used pre- and post-intervention, showed effects in terms of the Genevan model itself – the proportion of 14-year-olds showing early formal thinking was approximately twice that of comparable control schools. Large effects were shown 3 years later in performance on national General Certificate of Secondary Education (GCSE) assessments for science, maths and English (e.g. equivalent to a change relative to national proportions in English from 57.7% to 79.7% at C-grade and above). The norms were used in the teaching, in the assessment, and, it is claimed, were implicit in the better learning of the pupils as shown by their GCSE results (Adye & Shayer, 1990; Shayer, 1999; Shayer & Adey, 1993). Likewise, a cognitive intervention programme set within mathematics, CAME,9 has comparable transfer effects to GCSE science and English (Shayer & Adhami, 2007a).

More recently, a 1-year intervention for Year 1 children in Primary school, CASE@KS1 (Adye, Robertson, & Venville, 2002)10, featuring Piagetian concrete operational norms, and a 2-year intervention mainly in the context of mathematics, RCPCM (Shayer & Adhami, 2007b)11 covering Year 1 and Year 2, have both reported gains in KS1 maths and a doubling of the proportion of 7-year-olds at the mature concrete level on the Piagetian test of Spatial Relations, when compared with national norms. Much of RCPCM work was made possible by the research on the mathematics of young children by Terezinha Nunes with acknowledged debt to Piaget (Nunes, 2002; Nunes & Bryant, 1996). Nunes et al. (2007) report that, as part of an intervention study, 6-year-old children were given four subtests of the British Ability Scales, a test of working memory, and a four subtest battery of Piagetian logic tasks. The predicted variable was their Key Stage 1 SAT Maths scores. After entering first the BAS scores into the regression equation, followed by working memory (just significant, \( p = .04 \)), the Logical Competence test still predicted further variance, significant at the .001 level. When entered first into the regression equation, Logical Competence predicted 44% of the variance in SAT levels. Nunes concluded that ‘. . . children’s logical competence at the beginning of their school career is a significant predictor of their mathematics learning a year and a half later and a strong candidate for a causal factor of this learning’.

For all of this intervention work, the ‘Minute Particulars’ as reported in the Phase 2 Genevan studies were taken as valid descriptions of the path – sometimes branching – from non-possession of a norm to its achievement (Shayer & Adhami, 2003, give the maths context detail). Given the constructivist interpretation implicit in Piaget’s theory, the teacher realizes that to try to teach the norms would be a

---


10 CASE@KS1.H&F. Research project based at King’s College, London funded as part of a Single Regeneration Budget granted to the Hammersmith and Fulham LEA by the Government Office for London.

wasted effort. Instead, she has to manage the collaborative work of the children, only intervening to draw their attention to some aspect of the context they may not have considered in their work (she may, and does, remind them of their thinking in relevant contexts of their other NC lessons). Thus, the success of the interventions strengthens confidence in the validity of the underlying theory, since the application embodies the theory.

The INRC group and Piaget’s symbolic logic theory (Phase 2b)

With Piaget’s logical model – only prominent during Phase 2 of his life-work – there is a problem. Although it is quite informative in his description of the restricted logic of the concrete operations period – using set theory – the same cannot be said of its extension into formal operations. Piaget argued, in GLT, that propositional logic, in his version, can act as an ‘algebra’ for describing the structure of thinking, analogous to the use of algebra in physics. Perhaps it does describe a necessary aspect of formal thought. But it is certainly not the whole story. The pendulum task (GLT, chapter 4) was scaled as Class Task III as part of the CSMS work (Shayer, 1978), with the Exclusion of Irrelevant Variables items scaling at the mature formal (3B) level and the Control of Variables items scaling at the early formal (3A) level. Bond and Fox (2001, chapter 5) developed a pencil and paper test that modelled each of 16 binary operations of propositional logic in realistic contexts, as was done in Piaget (1949, pp. 224–265). Here is the item testing Implication.

A prospector has found that some rich metals are sometimes found together. In his life he has sometimes found gold and silver together; sometimes he has found silver by itself; every other time he has found neither silver nor gold. Which of the following has been found true for this prospector?

(a) Gold and silver are found together, never apart.
(b) If he found silver then he found gold with it
(c) If he found gold then he found silver with it
(d) If he found gold then he didn’t find silver.

It may be objected that this is a much easier version of implication than the abstract AD47 test pioneered by Peter Wason (Wason & Johnson-Laird, 1972, chapter 13), but it is Piaget’s modelling of it, which was fully contextualized, despite being offered in the clothes of propositional logic, viz Piaget (2006, p. 5):

Such a logic (Piaget, 1949) is already for the most part a logic of meanings but only with reference to the interpropositional structures.

Bond gave 150 14-year-olds both the BLOT 35-item test and also the CSMS Pendulum test, PRT III, and found that all items scaled together well under Rasch analysis. However, all the BLOT items scaled below all the Pendulum items except the mature concrete (2B) item for length, and even Implication scaled only at 6.8, just below the early formal (7+) level. At best this supports the hypothesis that competence in the 16 binary operations is a fairly easy necessary condition for success on the Pendulum task. Most of the binary operations scaled at the mature concrete (2B) or concrete generalization level (2B+).

In fact, Inhelder’s description of children’s performance in chapter 4 of GLT sufficiently exemplifies – as do all the earlier Genevan works – the norms of thinking that govern
the children’s actions, here being those that pertain to the control of variables and the exclusion of irrelevant variables. Piaget’s attempted re-descriptive explanation of them in terms of propositional logic is unconvincing.12

**Piaget’s Phase 3 work**

It is first necessary to insist that the whole of Piaget’s work has focused on the thinking of the ‘l’homme dans le rue’: the basic thinking that all humans share. He never studied their use of technical language, and designed the experimental contexts to avoid that. Nevertheless, there are a further set of more abstract norms — certainly in his mind during Phase 1, and to form the bulk of his work in Phase 3 — that historically have been one of the particular concerns of the philosopher.

‘... the emergence of logical necessity constitutes the central problem in the psychogenesis of logical structures.’ (Piaget, 1967, p. 391).

‘... now necessities are not observables. Rather, necessities intervening in reasons are formative with their links leading to that reconstitution of the object or event to be understood. ‘Reason’ consists in this.’ (Piaget, 2006, p. 8).

If the whole of Phase 2 is regarded as a travel around a wide circuit of the spectrum of performance norms — ‘performance’ in the sense that the concrete operational norms enable children to describe the world in order to operate on it, and the formal operational norms give access to the world of science and adult work — then in Phase 3 the norms of truth and necessity can be seen to underpin children’s use and application of the performance norms.

There are no less than 15 published books from Phase 3 of the Piaget’s work, so here only two which bear more directly on developmental psychology, affording possibilities for educational interventions, will be mentioned. In the last (Piaget & Garcia, 1987), Garcia quotes Piaget as having said ‘my logic needs cleaning up’.13 Piaget himself wrote (p. 11)

>The main aim of this work is to fill out and emend our operatory logic along the path of a logic of meanings.14

In this work he undertook to explore the further meaning of implication over and above its treatment in symbolic logic, from a realization of its close connection with necessary reasoning (Quine, 1952, pp. xv–xvii, 37, 38). Again, the strength of this work is the Genevan experimental method: the ability to create contexts that embody, implicitly, each aspect of thinking Piaget wished studied. The first study concerns the implications between actions as recorded on 1- to 4-year-olds. The apparatus related to that used by Köhler in his study of tool-use by chimpanzees. The out-of-reach three toy animals needed, respectively, the actions of lifting, drawing towards, pushing away and displacing to one side, to bring them within reach. ‘Implication’ here involves the child in such actions as ‘If *this* tool, then *that* action followed by . . .’ so the meanings were to do with the use of the eight possible tools provided and the situations of the animals.

---

12 The author has found no use whatsoever for this aspect of Piaget’s theory in any of his applied research.
13 ‘Il faut nettoyer ma logique.’
14 ‘Le principal but de cet ouvrage est de completer and de corriger notre logique opérateur dans le sens d’une logique de significations.’
The second study presents 5- to 12-year-olds with a tree-structure embodied in hollow opaque tubes leading in the end to eight possible destination boxes from a single tube beginning followed by two successive branchings. Here the norms studied were those of classification. Each of the tubes in the structure had a window that could be opened, and the child could see a coloured ribbon entering at the start. The questions asked were, ‘Without looking into the 8 boxes at the end, how many windows do you need to open to say into which box the ribbon goes?’ Here the meanings were to do with positions in the tree and the ‘if–then’ implications were to do with, having opened a window and seen the ribbon, how does that narrow down the search?

The third study, with 6- to 12-year-olds, involved tiling with squares, hexagons, pentagons, and triangles, the task being to select combinations of tiles which would completely fill a surface. The meanings are expressed in the properties of the shapes, and the implications are those of ‘if I first choose this, then which other(s) will do the job?’

The fourth study asks what happens if the issue is switched to the more abstract context of the child’s mental models of numbers. The study revisits the relations between cardination and ordination in 5- to 9-year-olds. A high box and a low box are connected with a tube. If, say, there are initially 11 balls in the high box, the question is asked, ‘Just before you let the 5th ball go, how many balls are there in each box?’ The ‘if–then’ is expressed in the questions, and solved in terms of the implication between the meanings of the cardinal and ordinal properties of number.

There are five more equally interesting studies, including a delightful study of weaving with different colours in both weft and woof conducted by Pieraut-Le Bonniec. In a brief conclusion Piaget admits (p. 145), implicitly, that his attempt in the *Traité de Logique* (Piaget, 1949) to make propositional logic do all the necessary work of explaining thinking was inadequate:

\[ \ldots \text{all acts of knowing include an inferential aspect, however implicit or elementary that may be} \ldots \text{all use of meanings assume and carry with them implications.} \]

Perhaps the attempt to logicize these findings is unnecessary: the varied details of the nine studies themselves constitute epistemological facts which are also a contribution to psychology.

*Studies in Reflecting Abstraction* (Piaget, 2001) contains an equally varied set of studies, almost exclusively confined to 5- to 12-year-olds; here only those possible for direct use for intervention purposes in school will be mentioned. Chapter 1 has Szeminska going over again ground earlier researched by her, but this time looking at how young children integrate notions of measurement, length, and number. Chapter 2 looks at ways in which children develop and differentiate multiplicative relations from the starting-point of additive relations. Chapter 3 studies how children develop the ability to invert additive and multiplicative operations. Chapter 6 on the formation of analogies, the most wide-ranging in content of the studies in this book, aims to exhibit the ‘Minute Particulars’ corresponding to Spearman’s ‘eduction of correlates’.

The experimental context is ingenious: drawings of specifics – a feather, a dog, a boat tiller, a vacuum cleaner, a boat and so forth – were given, jumbled up, to children from age 5 to nearly 11. They were then asked to ‘put together the drawings that seem to go well together’. After the children had sorted them into pairs, for example, boat with
tiller, they first would be asked why the pairs went together. Then, in various ways
depending on the children’s first explanation, the focus was directed on how certain
pairs of pairs might still be considered as having analogical connection.

This English version is noteworthy through the contribution of the translator.
Campbell’s superb and detailed scholarship, with comments, explanatory footnotes,
cross-references to many of Piaget’s works and lengthy introduction makes the English
version twice, for the reader, what it was in the French original.

While the different levels of performance described in the chapters may be
considered as ‘well-researched facts’ to be reflected upon, Piaget in this book also
endeavours to apply a higher level abstraction to the data: ‘empirical abstraction’;
‘reflecting abstraction’; ‘pseudo-empirical abstraction’; ‘reflected abstraction’ and
‘metareflection’. The intention of these is to give qualitative descriptive power to
processes that lead from one level to another. What is dropped is the technical term
‘accommodation’. The following quote gives the flavour of the explanatory model
offered (p. 151):

Classes in extension are generated by reflecting abstraction: Reflection, following
projection on to a new plane of assimilation, draws out of intensional relations those
relations that determine extension.

Opinions will differ on the utility of these descriptions: the author’s opinion is that they
suggest a level of abstraction much more potentially useful than Piaget’s attempt, in
Phase 2, to apply symbolic logic to the GLT data, but that they need clarification and
differentiation by other researchers’ use in educational contexts.

In these, and other works of the period from 1965 on, Phase 3 moves on from the
sub-research programme that culminated in GLT, and puts fresh life into the overall
research programme that was Piaget’s contribution to the psychology of intelligence.
Instead of the stuttering attempts to define intelligence verbally that one sees on just one
page in Carroll (1993, p. 36), one has empirically researched studies that produce
descriptions as rich in detailed developmental data and technical terms as the
complementary account of the relationships between Mental Abilities delivered through
the practice of psychometrics in Carroll (1993) and Cattell (1971). If you wish to focus
attention on what intelligence is – rather than how its applications to various aspects of
reality relate to each other – then this is the research programme that not only supplies
you with insight, but also still has life in it for further research, as shown in Smith (2002).
Starting with an Inhelder and Piaget (1963) study (Itération et Récurrence) using the
Genevan method of individual interview on a sample of 100 Year 1 and Year 2 children,
he tested three hypotheses related to differences in performance in school Year in:

(a) children’s reasoning by mathematical induction based on iterated action
(b) children’s modal reasoning based on logical necessity
(c) children knowing how to count without knowing when to count

Although the study does, correctly, include statistical data about the differences
between Year 1 and Year 2 responses, the main strength of the study, as with Geneva,
lies in the great details of children’s talk related to a complex descriptive analysis of the
children’s reasoning.

If Smith’s study comes closer to developmental psychology (DP), Piérou-Le Bonniec
(1980) chose to extend the DE path of Phase 3 Genevan research, although still with DP
overlaps. Rejecting Piaget’s propositional logic and INRC group as inadequate to model
natural reasoning, she argued that modal logic better serves to describe the relations between possibility and necessity. From experiments with children between the ages of 5 and 10 she shows the development of a natural process of reasoning – that allows some ability to test reality by gathering information – in terms of the three alternatives: necessarily true (or false); undecidable; and possible (p. 153). The achievement of this system by some 10-year-olds suggests that it corresponds to Piaget's concrete generalization stage, which she suggests (p. 58) explains more readily the swift acquisition of formal thought at the age of adolescence, the children being already half-way there. However, the details of this study show a third Euler circle – linguistics – overlapping those of DE and DP, and this, indeed was the direction her subsequent work has taken.

Demetriou: The neo-Piagetian criterion-referenced approach
A key evidence-based source of this approach derives from work carried out in the early 1980s, but only published after prolonged fights with editors and referees. Shayer, Demetriou, and Pervez (1988), reporting replications of Piagetian research on children aged 5 to 10 years, gave population estimates for a wide spectrum of different Piagetian tasks, based on data from UK, Greece, Pakistan, and Australia. The same wide spectra of development reported for the secondary students from the CSMS survey were found with the younger children. However, this also included the information that in each of the countries about 20% of the 7-year-old children were already at the mature concrete level on at least two-thirds of the tasks. In addition, it was shown that much of the ‘décalage problem’ – that children may show different levels of processing in different tasks - could be explained by a mental abilities model differentiating between classification and seriation tasks, conservations, and spatial tasks. More importantly, the Longeot (1978) hypothesis that there were nodes in development was confirmed by the data. Although between one node and another representing different qualitative levels of processing children differ, within limits, on different mental ability tasks, they do not proceed above the next higher node in one ability without having caught up on all the others at the preceding level.

There are two quite different meanings of the word ‘hierarchical’ in use. From psychometrics, by abstraction from the variance of the correlated mental abilities it is possible to define and measure a variable or variables (Carroll, 1993, Three level model) that expresses what is common to all. From Geneva, the hierarchy is of qualitatively higher and higher cognitive levels, with mature formal operational at the top. From the late 1980s, Demetriou (Demetriou & Efklides, 1994, p. 77) used both in his construction of test materials:

... our theory is inspired by the assumption that the developing mind can only be understood if the strong points of the developmental, the psychometric, and the cognitive tradition are allowed to converge and become integrated into a comprehensive system.

Perhaps the Lakatos (1978) rendering of the progress of science, confined mainly to that of physics, may need further refinement when applied to psychology. Yet the psychometric and the Genevan research programmes, each excluding the other as their negative heuristic, and each with a strong positive heuristic, are in fact describable in his terms; and he did say, in objection to Kuhn (p. 69):

The history of science has been and should be a history of competing programmes (or, if you wish ‘paradigms’), but it has not been and must not become a succession of periods of normal science: the sooner that competition starts, the better for progress.
He even said (p. 69) “Theoretical pluralism” is better than “theoretical monism”

These might have been for him good examples to quote. Here the complementarity metaphor begins to break down, even if it did describe 30 years work in mutual exclusion: if Demetriou’s methodology is valid then it is possible to combine the two research programmes to the benefit of work on ‘intelligence’ without losing measurement information.

Demetriou and Efklides (1994), using distinctions drawn from the mental abilities literature, produced test batteries of ‘specialized structural systems’ (SSSs): qualitative-analytical, quantitative-relational, causal-experimental, verbal-propositional, and spatial-imaginal. But he used test-theory in a novel manner: by using a rectangular distribution of item-difficulty and Rasch analysis or discrimination analysis (Demetriou, Efklides, & Platsidou, 1993; Demetriou & Kyriakides, 2006), he built into the programme the hierarchy of cognitive levels derived from and described in the Genevan research (in this aspect from Phase 2 work). But as part of the research he used also ‘... clinical interviews and longitudinal examination, or experimental methods, such as training experiments, to cross-validate hypotheses about structure and development’ (Demetriou & Efklides, 1994, p. 76). From the cognitive realm he used tests of speed of processing, control of processing, and working memory. Finally he included tests of ‘the hypercognitive system’ – that is tests concerning children’s own thinking about their own thinking.

In Demetriou (2004, Table 1.1, p. 42–44) two products of the programme are presented: firstly a detailed taxonomy of children’s competence at all ages – as in Binet (1908; Binet & Simon, 1980) many years earlier – from 3/4 years to 15/16 years on all five of the SSSs: that is, a hierarchical description in the Genevan sense. In addition, by using the extension of the factor-analytic programme – structural equation modelling – he was able to explore both the issue of general and specific factors – the hierarchy in the psychometric sense – and also their relation to age and to the experimental psychology measures of speed of processing, control of processing, executive, phonological, and visual memory. In this way the measured variables of cognition were provided with a research methodology exploring some causal mechanisms of development related to research of brain development.

**Can neuroscience contribute to the concept of intelligence?**

Perhaps one can apply the metaphor of describing, separately, the skill and mental world of a Formula One driver, and the mechanisms of his car supported by the technical team: the former being analogous to psychology and the latter to neuroscience. Duncan *et al.* (2000), using two tests derived from the normal science of psychometrics, each with high g content, sought to distinguish between the Spearman and the Thomson account of brain functioning (the distinction between reading the results of factor analysis as corresponding to specific locations in the brain, and processing being spread over the entire brain). One test was in the spatial mode; the other in the verbal. Participants were given PET scans while working on the tasks. In each case areas of the cortex associated with spatial or verbal content showed increased blood flow, indicating activity, these severally being in quite different areas of the cortex. But common to both was activity in the same area of the prefrontal cortex, an area associated with executive control and strategy (functions represented by tests used by Demetriou as described above). Duncan

---

16 This question is blessed in Piaget (1995), Sociological Studies, p. 23, para 2.
argued that this was evidence for a specific neural basis for $g$. Duncan et al. (in press) report further research in this line bearing on factors that can interfere with processing, Houdé and Tzourio-Mazoyer (2003) cited evidence given on adults working on an implication task such as modelled originally by Peter Wason. This was made slightly harder by being in the form ‘if not $p$, then $q$’ – ‘if there is not a red square on the left, then there is a yellow circle on the right’ – and the failure rate was 90%. That is, on being asked to choose objects which would make the rule false, they put a red square on the left of a yellow circle. Houdé & Tzourio-Mazoyer’s argument was that people have two competing strategies in their mental workspace – one perceptual and the other logical – but they have trouble inhibiting the perceptual one. The perceptual strategy leads to their using the shapes that are mentioned in the rule rather than to see whether a blue square on the left gave, say, a green shape on the right. Giving the participants’ logical explanations or just more practice examples made little difference to their failure rate, but when they were given emotional warnings about the error risk if they did not specifically inhibit their intuitive perceptual strategy, then their success rate became 90%. Hence, it was not their logic that caused their initial error. fMRI imaging showed that it was the right prefrontal cortex as well as the left (these are the brain areas to do with control strategies) that was involved in the cognitive inhibition, with the implication that emotion probably assists reasoning (it can also interfere with it!).

This suggests that in cognitive development processes very often perceptual inhibition may be a necessary condition. For example, in Piaget’s classical 1:1 correspondence problem children will not succeed unless they inhibit their perception that the length or the space occupied by one of the set of six counters affects their numerosity. Houdé & Tzourio-Mazoyer’s comment on this is:

Developmental studies also indicate that children often fail to inhibit reasoning biases, especially perceptual ones (or semantic belief biases), and that they are even more receptive to bias-inhibition training than are adults.

Shayer and Adhami (2003) report considerable detail of sources used in designing lessons for the RCPM project for Year 1 and Year 2 children in an intervention designed to increase their intelligence, and cites use both of the above, of Butterworth (1999) on the Number Module, and Dehaene (1997; Dehaene, Dehaene-Lambertz, & Cohen, 1998) on children’s apprehension of magnitudes.

In terms of the Formula 1 metaphor, just as it would be foolish to ask a driver to go out in a car with faulty steering, so too one cannot ignore evidence on what areas of activity intelligence can operate well, particularly in the intervention studies previously mentioned.

**The Flynn effect**

Flynn (1987, 1994), surveying data on the standardization and re-standardization of various psychometric tests from the late 1940s on, reported gains of about 9 points per generation for tests of crystallized intelligence, and 15 points or one standard deviation for tests of fluid intelligence (Raven’s matrices). This general apparent improvement has
entered the literature as ‘the Flynn effect’. In a later review, Flynn (1998), examining evidence going back to the 19th century, had become sceptical of claims that this meant that people’s intelligence - children or adults - was improving generation by generation. He wondered whether the concept of ‘intelligence’ was still viable. Certainly, people’s test-taking ability has improved; but after discounting as minor factors such as nutrition, urbanization, and TV and pointing out that children’s gains on arithmetic reasoning, vocabulary, creativity and speed of learning were far less, he was tempted to suggest that school and environmental stimulus in some undefined way were improving children’s decontextualized problem-solving skills.

Yet, this may be only a historic effect; Grissmer, Williamson, Kirby, and Berends (1998, pp. 262–265) argue that there is some evidence suggesting that gains have been levelling off from the late 1980s. In this country, the NFER-Nelson CAT test showed no change in norms between the 1984 and 2000 standardizations. Flynn (2007) wonders whether IQ gains will cease in all highly industrialized nations, citing Sundet, Barlaug, and Torjussen (2004) as evidence that gains have ceased in Norway; and Emanuelsson, Reuterberg, and Swenson (1993) and Teasdale and Owen (2000) that gains have flattened since the late 1980s in Sweden and Denmark, respectively. Flynn argues that in the Scandinavian countries the whole population has by now adjusted to 20th century industrialization and its implications.

Yet Flynn (2007) also shows, when the subtests of arithmetic, information, and vocabulary in the WISC are examined, between 1947 and 2002 children gained only an average of 3 IQ points when compared with a change on Raven’s of 27.5 points. If ‘real intelligence’ is that which enables school learning, children’s intelligence has hardly changed at all over 55 years. An alternative explanation that the schools have got that much worse in teaching the 3Rs seems hardly likely.

The Flynn effect is a major scandal for the psychometric programme – while 50 years of sophisticated work of measurement was going on in the world of psychometrics their ruler was made of rubber, and stretched over a standard deviation (nearly two on tests of fluid intelligence). Moreover, different measures have been sliding by different proportions.

Contrary evidence in Shayer et al. (2007) on some 10,000 Year 7 pupils on the criterion-referenced CSMS Piagetian test Volume & Heaviness - a test of conservations and density - showed that between 1975 and 2003 there had been a drop in performance of 1.04 SD for boys, and 0.55 SD for girls, and that an initial boy-girl differential of 0.5 SD had completely disappeared. If Flynn is right in saying that the change in the general environment is responsible for the gains in decontextualized problem-solving skills for psychometric tests, then the general environment change has been very unfavourable for boys (and not all that favourable for girls either) on Piagetian measures, at least those concerning conservations and density. Both boys and girls were down, also, on the 3 items testing early formal performance on volume/mass relations.

Flynn (2007) also cites British evidence that from 1990 to 2003, British children lost 0.4 SDs on the WISC Arithmetic subtest – a rate, he argued, which is identical to the rate of decline shown in Shayer et al. (2007) for the Piagetian conservations.

Evidence on the ‘plasticity’ of intelligence
In a lengthy survey of the psychometric, information processing, and Piagetian literature Adey, Demetriou, Csapo, Hautamaki, and Shayer (2007) argued that the evidence
supported a concept of Plastic General Ability underlying intelligence (‘Plastic’ signifying modifiable, not fixed). For ‘Plastic’, the historical document Dasen (1972) showed that the variation in appearance of formal operations was explicable in terms of the selective urbanization of rural West Africans, and also the Flynn effect literature was cited. For ‘General’ the evidence on g (or g values) from the work of Cattell and Carrol, was referred to, with the reminder from Gustaffson (2002) that the wider the spectrum of psychometric tests used, the stronger the estimate of g.

Given that both psychometrics and the Duncan study from neuroscience support the ‘General’ part of that argument, it is to intervention studies that one must look for other evidence for the ‘Plastic’ part. Feuerstein’s Instrumental Enrichment (Feuerstein, Rand, Hoffman, & Miller, 1980) programme (FIE) is a two-year school intervention, typically initiated at the beginning of adolescence, designed to increase children’s ability (and belief in that ability) to benefit from exposure to challenging thinking tasks that cover the spectrum of schemata described by Piaget and some from the mental abilities literature. Collaborative learning – that is peer–peer mediation – is an important feature of FIE practice. Shayer and Beasley (1987) reported on the use of FIE with one class in a Special School in UK, where effect-sizes of 1.22 SD on a battery of Piagetian tests, and 1.07 SD on Raven’s Matrices were achieved – in IQ terms this is equivalent to change from a mean of 85 to a mean of 103. Arbitman-Smith, Haywood, and Bransford (1994) reporting replication of FIE in the USA, cited effect-sizes of the order of 0.6 SD on Raven’s and other tests of fluid intelligence, averaging 0.77 SD on tests of crystallized intelligence. These data – together with the CASE, CAME, and RCPCM evidence given earlier – challenge the notion that the IQ of a child will remain constant from year to year even though their general learning is progressing together with their contemporaries.

Summary
The switch to a criterion-referenced methodology by Demetriou and others made it possible, among other questions, to ask and describe what qualitative changes were present when children’s thinking rises to quantitatively superior levels. Hence, the work of Demetriou in particular amounts to a research programme, including the strengths of psychometrics, which still may have some life in it. For example, Demetriou has already collaborated with the philosopher Kargoulos to address the relationship between propositional logic and three of the SSSs (Kargopoulos & Demetriou, 1998). Although the CASE, CAME, and RCPCM interventions were informed both by Piagetian and Vygotskian theory (Shayer, 2003) the evidence gathered bore only on the magnitude of the cognitive gains that could be achieved by large samples of children. Further research should be possible on selected samples, using the Demetriou measures of better management of mental resources (processing variables), and his tests of the hypercognitive systems, to investigate what accompanies the cognitive gains.

The conception of intelligence presented in this paper is one that informs a research programme directed towards the better experience of children in school, of which it can only claimed that the first words have been written. The ‘hard core’ of this programme is closely related to the equilibration negative heuristic of Geneva: that is, from a consideration of the totality of the interaction of children in their environment – partly derived from good ‘Minute Particulars’ research and partly developed by teachers’ creative action in the classroom (Shayer & Adhami, 2003) – teachers’ practice can be informed by an intuitive sense of how to use the collaborative learning of children in the
service of their construction of their ever more powerful learning and reality-processing
in the world. Smith (2002, chapter 8) gives a good account of how much Piaget's own
interest in education anticipated such practice. But further development must be backed
by good applied and applicable research, which both develops the art, respecting
teachers' own professional autonomy, but also - unlike much Government practice -
continually looks for validation to good evidence-based research.

Acknowledgements

Substantial assistance to thought was provided initially by a long dialogue with Les Smith and later
by the Editor, Andrew Tolmie. However, neither is to be held responsible for the argument
presented by the author.

References

intelligence? Why education needs the concept of plastic general ability. Educational
1 pupils. British Journal of Educational Psychology, 72, 1–25.
Brooks, R. Sperber, & C. McCauley (Eds.), Learning and cognition in the mentally retarded
Williams and Wilkins.
Bond, T., & Fox, C. M. (2001). Applying the Rasch model: Fundamental measurement in the
Psychology, 3(1), 23–40.
(Eds.), Cognitive developmental change: Theories, models and measurement (pp. 21–73).
Cambridge: Cambridge University Press.
mind: Experiential structuralism as a frame for unifying cognitive developmental theories.
Demetriou, A., & Kyriakides, L. (2006). The functional and developmental organization of


