A Dynamic Test to Assess Learning Capacity in People With Severe Impairments

Christine Hessels-Schlatter
University of Geneva

Abstract
Because traditional intelligence tests do not allow a reliable and valid estimate of the general cognitive abilities of individuals with IQs lower than 55, a dynamic test of analogical reasoning (Analogical Reasoning Learning Test—ARLT) was constructed for use with this population. The aim of the test is to distinguish individuals who could undergo more ambitious schooling and cognitive training from those who would profit little or none from such an intervention. Internal consistency and test-retest reliability of this measure are high. Discriminant and predictive validity, assessed with another learning test and with a one-month training on inductive reasoning, respectively, are satisfactory. The test is seen as a useful tool for diagnostic differentiation among low performers.
uals with similar types or degrees of mental retardation diverge significantly in their lifelong development (Landesman Ramey, Dossett, & echols, 1997). In social ecology (Berkson & Landesman-Dwyer, 1977; Landesman Ramey et al., 1997), the assumption is that individuals directly influence and are influenced by their social environments.

Individuals with moderate or severe mental retardation are generally considered as one group and receive the same schooling, although their etiology as well as their learning experiences are different. Furthermore, some researchers (e.g., Büchel, Schlatter, & Scharnhorst, 1997; Budoff & Hamilton, 1976; Tzuriel & Klein, 1985) have suggested that the learning capacities of these individuals differ as well. Yet, it is very difficult to make an estimation of this potential because traditional intelligence tests are considered to be irrelevant for this population (Burns, 1990; Hessels-Schlatter, 2002; Kamphaus, Kaufman, & Harrison, 1990; Reschly, 1992; Schlatter & Büchel, 2000). Several characteristics of the population (e.g., lack of understanding of task instructions and demands, short attention span, weak communication skills, and slow information processing) lead to a general floor effect on traditional intelligence tests. Thus, the results are neither reliable nor valid. With regard to reliability, Wishart and Duffy (1990), for example, have clearly shown children with Down syndrome do not show stable performances on intelligence tests. Süß-Burghart (1995), in a study that included children with moderate to severe mental retardation, demonstrated that a differential diagnostic with the Kaufman Assessment Battery for Children is not possible for this population. Empirically, the norm tables do not include enough (or have no) individuals with an IQ below 60. Reschly (1992) stated that because none of the current intelligence test norm tables go below an IQ of 40 or 50, IQs lower than 55 must be extrapolated considerably. In these circumstances, the definition of moderate, severe, and profound mental retardation is not possible: “The IQs below 55 necessary to operationalize the AAMR and APA [American Psychiatric Association] criteria for MR [mental retardation] levels are more illusion than reality!” (p. 41). The validity problem arising with the use of traditional intelligence tests with this population is exemplified by Carlson and Wiedl (2000) as follows:

If a student fails to solve an arithmetic problem such as “Mary has 8 apples and Peter brings her 5 more, how many apples will Mary have?” the conclusion usually would be that he or she cannot add 8 plus 5:

- But how sure can we be about our conclusion? Perhaps the student knows what 8 + 5 is but does not recognize that the addition operation is required for this word problem. Or perhaps the student cannot read the question and, unknowingly, we have given the student a reading test rather than a math test. Or perhaps the student selected an answer randomly, never reading the question. Again, unknowingly, we may have tested the student’s attitude, motivation, ability to follow directions to comply with our expectations or whatever. The point is that for a variety of reasons the student’s performance may or may not reflect what we think it does and the conclusions we draw. Accordingly, the inferences made may be inaccurate and misleading, and we have a validity problem. (p. 681)

Furthermore, intelligence tests provide a static image of general cognitive abilities, which are assumed not to change over time. In this perspective, educational intervention consists mainly of changing the environment (i.e., to place the student in a less demanding setting). However, planning intervention should include consideration of the potential gains that individuals may acquire through intervention. A direct consequence of the label of moderate or severe mental retardation, or of practically educable, is the limitation of learning and social experiences necessary for development (Feuerstein, Rand, & Hoffman, 1979; Smith, 1997; Taylor, 1997). Campbell and Carlson (1995) stated that “a child who appears retarded... may be trained to be retarded because of limiting educational opportunities” (p. 12). Finally, the absence of a relationship between tests scores and educational intervention has been expressed by Glaser (1981), who emphasized that it is perhaps the schools that must be flexible enough to offer differentiated schooling according to the students’ needs.

In the area of cognitive abilities, few studies have been focused on the moderately or severely mentally retarded population. Most research has been centered on the acquisition of language and practical or social skills. Only in the last 10 years have researchers started to systematically study the strengths and deficiencies in information processing of, for example, individuals with Down, Williams, and fragile X syndromes (e.g., Hodapp & Dykens, 2001; Hodapp et al., 1992; Molina & Perez, 1993). These researchers, however, did not tackle the question of intellectual enhancement and possibilities of learning for individuals with moderate or severe mental retardation.

Dynamic procedures have been proposed as
Alternatives or supplemental procedures to traditional intelligence tests for special populations (Budoff & Friedman, 1964; Budoff & Hamilton, 1976; Carlson & Wiedl, 1978, 1992; Feuerstein et al., 1979; Guthke, 1992; Hessels, 2000; Stubbings & Martin, 1998; Swanson, 2000; Tzuriel, 1997). The Analogical Reasoning Learning Test—ARLT (Hessels-Schlatter, 2002; Schlatter, 1999; Schlatter & Büchel, 2000) was especially constructed for the assessment of individuals whose IQ, as measured with a traditional test, would be below 50–55. The aim of this test is two-fold: (a) to provide a reliable and valid estimation of the general cognitive abilities and learning capacities of individuals with moderate or severe mental retardation and (b) to explore whether some of these individuals can achieve a level of abstract reasoning and learn more than is usually expected. On a practical level, the ARLT should allow the planning of differentiated, and for some individuals, more demanding schooling according to their strengths.

The ARLT is a dynamic procedure intended to assess learning capacity of children, adolescents, and young adults with moderate to severe mental retardation. Although the test can, in principle, be administered to children of all ages, use of the test would be most appropriate in the age range of about 8 to 12 years. The aim of the test is to distinguish persons with moderate or severe mental retardation who can profit from cognitive training programs and more demanding schooling from those for whom such an approach would have little value. A number of arrangements with regard to test material and administration procedures were developed to avoid the difficulties encountered when employing classical tests. The presentation format of the test should permit quick understanding of both the problem to be solved and the associated instructions. In order to teach the students the basic principles and rules to be applied, the examiner includes a preparatory phase before testing. A good mastery of language is not required. The examiner provides a description of the task as an external support in the problem-solving process.

The ARLT consists of 2 × 2 analogical matrices, mostly in a figurative concrete modality (see Figure 1) and some in a geometrical modality. Two levels of complexity are distinguished. At the first complexity level, there is one relation between the Elements A and B and another between A and C. At the second level, the items require the induction of two relations between the Elements A and B and one between A and C. The elements of the matrix as well as possible responses are printed on 8 × 8 cm wooden squares. Each matrix is constructed by the examiner on a wooden frame, which contains four boxes. Once the first three elements (A, B, and C) are placed, the participant is asked to select the fourth element (D) from a set of 6, respectively 8 possible responses.

The ARLT is divided into three phases (see Figure 2). The first is a pretraining phase intended to familiarize students with the task demands and to teach them some cognitive prerequisites. With simplified analogies (Phase 1a), the students learn how to infer and apply one relation and, subsequently, two relations. In case students are unable to abstract differences between the elements, Phase 1b (comparison) is administered, whereby the students learn to systematically compare elements. The aim is not to ensure that students give a perfect verbal description but, rather, to lead...
them to compare task elements in a systematic way, to detect similarities and differences by distinguishing relevant from irrelevant aspects, and to communicate these (also nonverbally).

In the second phase, the students have to solve seven items at the first complexity level. The goal of this learning phase is to teach them to solve analogical matrices, with the help of specific, standardized, and hierarchically ordered hints. Two processing stages are distinguished: description and solution. The aim of the description stage is to focus the students’ attention on the matrix, to guide them precisely to encode the elements and to compare the elements before looking at the set of possible responses. The set of multiple choice responses is then placed in front of the students (solution stage), and they are instructed to choose one response and place it in the empty box. If the selected answer is correct, positive feedback is given and the principles required for task solution are summarized. If the selected response is wrong, a standardized and error-specific sequence of hints is provided.

The third phase takes place one week after learning. This phase was designed to evaluate maintenance and transfer capacity of the learned rules and is applied in a more static way. After an introductory item, three maintenance items (similar to those of the learning phase) are given. No hints are provided in this third phase. If the students fail, appropriate feedback is given (“That’s incorrect,” or “That’s almost correct, try again”), and they can try a second time. The near transfer items consist of four new items at the first complexity level. For far transfer, students are presented with three items at the second level of complexity, after an introductory item.

The estimate of learning capacity is based on the maintenance and transfer phase. Indeed, the learning phase, with its dynamic interaction ought not to give a reliable measure, in contrast to the third phase. The students are in a learning process and, thus, great inter- and intraindividual variability should be expected. In contrast, the third phase permits an estimation of what the students have learned and are able to apply in a more independent way. Scores are assigned according to the need for help and the ability to solve the items. Items solved without feedback are credited 2 points. Items succeeded after negative feedback are credited 1 point, and items still failed after feedback are credited no points. The maximum score for the maintenance and transfer phase is 20.

Learning capacity is operationalized at three levels, corresponding to a status of gainer, nongainer, or undetermined. The terminology originally proposed by Budoff (1987) was adopted: The gainers are those who profit from the learning phase in the ARLT; the nongainers profit little or not at all. Next to these two extreme groups, an intermediate group, called undetermined, was distinguished, for whom other criteria should also be considered. Clearly, learning capacity ranges along a continuum and proposing a cut off score between gainer and nongainer is not wise. The present study was conducted to determine the reliability as well as predictive and construct validity of the test.

Method

Participants

The study involved 58 students (38 males, 20 females), all educated in special schools for students with moderate to severe mental retardation. Their mean chronological age was 13.92 years (range = 6.50 to 19.83, standard deviation [SD] = 3.0). There were records of IQs (mostly on the Wechsler Intelligence Scale for Children–III [WAIS–III] and the Kaufman Assessment Battery for Children) for only half of the students (n = 30). The scores ranged from 30 to 58 (M = 43, SD = 9.4). The mean age of the experimental group was 14.42 years (SD = 2.9); their mean IQ was 39.4 (SD = 8.8). The mean age of the control group was 13.50 years (SD = 3.2), and their mean IQ was 47.3 (SD = 9.0). Both groups consisted of 19 males and 10 females.
Table 1. Experimental Design

<table>
<thead>
<tr>
<th>Week</th>
<th>Design</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pretest: 1st part ARLT*: pretraining and learning</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>Interval</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>Pretest: 2nd part ARLT: maintenance, near and far transfer</td>
<td>—</td>
</tr>
<tr>
<td>4-6</td>
<td>Interval</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>Retest: 2nd part only: maintenance, near and far transfer</td>
<td>Test reliability</td>
</tr>
<tr>
<td>8-9</td>
<td>Interval</td>
<td>Concurrent validity</td>
</tr>
<tr>
<td>10</td>
<td>Raven Kurzzeit-Lerntest</td>
<td>Predictive validity</td>
</tr>
<tr>
<td>11-14</td>
<td>Training (experimental group only)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Posttest</td>
<td></td>
</tr>
</tbody>
</table>

*aAnalogical Reasoning Learning Test.

Design and Procedure

The experimental plan followed a pretest–retest–training–posttest design, as shown in Table 1. The complete ARLT was administered as a pretest to all students. After 4 weeks, the third testing phase (maintenance and transfer) was repeated in order to assess test–retest reliability. As discussed in the test presentation, we assume that the scores obtained in the maintenance and transfer phase are stable and reflect the student’s capacity to profit from the learning phase. In contrast, the dynamic procedure of the learning phase should lead to inter- and intraindividual variability, and, therefore, these scores will be unstable.

The students were matched on the basis of their ARLT performance (maintenance, near and far transfer) to evaluate the predictive validity of the test and were randomly assigned to either an experimental group with training in inductive reasoning or a control group without training. The training for the experimental group consisted of 8 to 12 lessons of about half an hour, at a rate of two or three lessons per week over a period of 4 weeks. Eight of these lessons were provided to all students; the other 4 were designed to individualize the training by providing prerequisite lessons to students who did not reach a fixed learning criterion. The students were trained in pairs in order to provide some social interaction during learning.

The lessons included different kinds of tasks and varied presentation formats. To begin with, standard analogies (involving similar kind of relations as in the ARLT) at two complexity levels were trained. Then, analogies requiring the induction and application of other kind of relations, such as “lives in,” “one as opposed to many,” and “is part of” were introduced. The presentation format of all these tasks varied (e.g., photographs, wooden blocks, educational games) in order to simulate, as far as possible, realistic everyday problems. Finally, classification tasks were also trained. After each kind of task training, evaluation tasks were administered, to observe individual learning progress and to distinguish the students that needed supplementary lessons. The evaluation tasks were thus seen as an essential part of the training.

With regard to the selection of validity criteria, it is believed that these should correspond to the task domains as well as to the style and quality of the instructions provided in learning tests. The training provided in this study was well-controlled with respect to content, declarative knowledge, complexity level, and instructional variables. The posttest was designed to evaluate the training effects.

All students were tested immediately after training. The items included in the posttest correspond to the training tasks (i.e., analogies involving the different types of relations mentioned above as well as classifications) in order to evaluate training effects. The presentation format was the same as in the training, but the examiner did not provide any hints. For each item where they had failed, the students could try a second time, after having received appropriate feedback (“This is incorrect” or “This is almost correct, try again”). A total of 28 tasks (13 standard analogies,
8 analogies defined by other relations, and 7 classifications) were administered. The scoring procedure was the same as the one adopted in the ARLT (i.e., 0, 1, or 2 points), leading to a maximum of 56 points.

The Raven Kurzzeit-Lerntest (Frohriep, 1978; Guthke, 1992) was also administered to estimate the concurrent validity of the ARLT. This measure is a learning test based on the Coloured Progressive Matrices (Raven, 1965), in which the test items are presented in a puzzle form together with a set of standardized and hierarchical hints. It was originally developed to detect developmental delay among kindergarten children who were likely to require special education.

Results

Reliability

The 10 maintenance and transfer items of the ARLT show a high internal consistency, with a Cronbach’s alpha of .88. The same coefficient was found for the retest. Test–retest reliability yielded a Pearson correlation of .83, which meets generally accepted reliability standards (see, e.g., Nunnally & Bernstein, 1994).

Validity

The validity data are based on different composite measures: (a) the posttest score on the analogy items, which included 13 standard analogies and 8 analogies defined by other relations (alpha reliability = .89); (b) the posttest score on the 7 classification items (alpha = .88); and (c) the posttest total score, which is the sum of all analogy and classification items (a total of 28 items), with a Cronbach’s alpha of .93.

Our sampling was rather broad with regard to age: Most of the participants were between 9.08 and 17.83 years of age; one student was 6.42 years old, 2 were 18, and one was 19.83. Nevertheless, in an analysis of variance (ANOVA) with group as the independent variable and age as a covariable, the age effect on the pretest (maintenance and transfer phase) proved to be nonsignificant.

Predictive Validity

Before estimating the predictive validity of the test, I had to show that the training effect, as reflected by the posttest measures, represented a valid criterion. For the experimental group, the true training effect was confounded with the test repetition effect, but the latter can be controlled for by means of the control group. The differences in performance between the two groups, therefore, express the pure training effect.

The ANOVAs with group as the independent variable showed that the differences between trained and untrained groups were significant for the posttest score on the analogy items, $F(1, 57) = 3.32, p < .05$ ($M_s$ of the experimental group and control group = 21.7 [SD = 11.8] and 17.3 [SD = 8.7], respectively), the classification items, $F(1, 57) = 9.07, p < .01$ ($M_s = 10$ [SD = 3.7] and 7.4 [SD = 4.8], respectively), as well as for the posttest total score, $F(1, 57) = 5.33, p < .05$ ($M_s = 31.7$ [SD = 14.9] and 24.5 [SD = 13.1], respectively). The effect size, measured with $d$ equals (mean of experimental group minus mean of control group divided by SD), was .50, which is sufficiently high to be considered as different from a simple familiarization effect (Cohen, 1988). The SDs proved to be rather large. This can be explained by the fact that both groups contained gainers and nongainers who were not yet distinguished in this analysis.

Furthermore, even though the students were matched with regard to the results on the pretest, age, and institution and were randomly assigned to one of the two groups, a significant institution effect existed on the three measures. Indeed, students came from five different institutions, and this result seems to demonstrate that the schooling in these institutions is rather different. However, there was no interaction effect between group and institution in the ANOVA, as expected.

Differential effects of the training with regard to status (gainer, nongainer) and group (experimental, control) was expected because the efficacy of the training has been proven; that is, an interaction for Status × Group. Individuals diagnosed as gainers in the test should have profited more from the training than did the nongainers. Also, one would anticipate a group difference: The gainers in the experimental group should get higher scores on the posttest than would the gainers in the control group because the latter were not trained. However, the scores of the gainers in the control group varied from pre- to posttest, reflecting a capacity to profit from test repetition without additional instruction. With respect to statistical treatment, this double hypothesis was tested using an ANOVA with group (experimental, control) and status (gainer, nongainer) as factors.
For the analogy items, the effects of group and status were significant, $F(1, 55) = 4.34, p < .05$ and $F(1, 55) = 166.83, p < .001$, respectively, as well as the interaction Group × Status, $F(1, 55) = 5.56, p < .05$. This interaction means that the gainers in the experimental group obtained significantly higher scores than did those in the three other groups (i.e., the nongainers in the experimental group, the gainers and nongainers in the control group) as predicted. Similar results were found for the classification items only, with a main effect for group, $F(1, 55) = 7.74, p < .01$, and status, $F(1, 55) = 84.70, p < .001$, and a significant interaction, $F(1, 55) = 3.34, p < .05$. However, the interaction effect in this case was in the opposite direction. This effect was mainly due to the fact that the gainers in the control group were able to rather easily solve classification items without specific training (see Table 2). Moreover, there was a ceiling effect for the gainers in the experimental group (who had a mean score of 13.1, with the maximum being 14) as well as for the gainers in the control group ($M = 12.4$). Unfortunately, we did not have sufficient classification items, particularly more difficult ones, to determine whether the gainers in the experimental group would have outperformed the gainers in the control group.

It can also be noticed in Table 2 that the nongainers in the experimental group showed a familiarization effect on the classification items after training because their performances were higher than those of the nongainers in the control group but that they were resistant to the training on the analogy items. With regard to analogy tasks, the gainers in the control group still obtained higher scores than did the nongainers. This is mainly due to the fact that the 10 items of the maintenance and transfer phase of the ARLT were repeated in the posttest. The difference obtained by the gainers was located on those 10 items (which is normal because success on these items is the criterion for the status of gainer), but not on the other analogy tasks (other kind of relations).

The data confirm the predictive validity of the ARLT. The students classified as nongainers on this test showed no improvement after training on inductive reasoning tasks and obtained the same results as the nongainers who were not trained. However, they were still able to profit a little from the training on the classification tasks. It is not surprising that the nongainers from the experimental group obtained higher scores than did the nongainers from the control group because the control group received these kinds of items for the first time during the posttest (familiarization effect). It was also interesting that the gainers in the control group outperformed the nongainers in the control group, especially on the classification items. We explain this by the fact that the pretraining and training phase of the ARLT gave them the tools to solve tasks that require the basic processes of inductive reasoning (i.e., systematic comparison of similarities and differences).

Predictive validity was also confirmed by the results of regression analyses, which show that for the experimental group, 65% of the variance in the posttest (the 10 repeated items excluded) is explained by the ARLT.

### Concurrent Validity

To estimate concurrent validity, I compared the scores obtained on the ARLT to those obtained on another dynamic test of analogical reasoning, the Raven Kurzzeit-Lerntest, which has been normed by Frohriep (1978) for children between 5.42 and 6.42 years of age. In ascertaining

### Table 2. Means and SDs on the Posttest Scores by Group and Status

<table>
<thead>
<tr>
<th>Group</th>
<th>Analogy items (max = 42)</th>
<th>Classification items (max = 14)</th>
<th>Posttest total score (max = 56)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gainers</td>
<td>Nongainers</td>
<td>Gainers</td>
</tr>
<tr>
<td>Experimental</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Experimental</td>
<td>34.7</td>
<td>5.6</td>
<td>10.8</td>
</tr>
<tr>
<td>Control</td>
<td>27.8</td>
<td>6.9</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>47.8</td>
<td>6.1</td>
<td>18.3</td>
</tr>
</tbody>
</table>

Note. $N = 10$ for the experimental group, 8 for the control group.
the reliability of the Raven Kurzzeit-Lerntest, I found a Cronbach’s alpha of .87.

The correlation between the ARLT score and the standardized score on the Raven Kurzzeit-Lerntest total score was .60. This correlation can be considered to be moderately high, suggesting that the tests partially measure the same abilities but also provide different information (Anastasi, 1997). Indeed, the Raven Kurzzeit-Lerntest consists not only of analogies but also includes continuous and discrete pattern completion tasks, which engage perceptual rather than reasoning processes. These kind of tasks are presented mainly in the first part of the test (Set A, \( r = .37, p < .01 \)) and rather less in the second part (Set AB, \( r = .53, p < .001 \)). The correlation with Set B, which consists of items that are mainly real analogies that cannot be solved perceptually, is the highest, \( r = .62, p < .001 \). The correlations with each of the three parts of the Raven Kurzzeit-Lerntest reflect the importance of analogical items. These differential correlations can be interpreted as a sign of discriminant validity as required by Anastasi (1997). This confirms the analogical nature of the ARLT.

Discussion

Evaluation of general cognitive abilities of students with moderate or severe mental retardation appears to be rather difficult and often useless with classical intelligence tests. However, even if the level of cognitive functioning of such students is very low, they cannot be considered as a homogeneous group, and more discriminative estimation of their abilities is needed in order to plan a proper educational curriculum with regard to their learning capacities. The ARLT has proven to be a reliable and valid instrument for the estimation of learning capacity with regard to analogical reasoning in individuals with moderate or severe mental retardation. As stated, learning capacity is a rather stable characteristic that can be assessed through a dynamic procedure. The predictive power of the test is very satisfying, and the fact that the gainers in the test also are the gainers in the training shows that a link can be established between a cognitive component and instruction, which reinforces the conceptual validity of the test (Carlson & Wiedl, 2000). However, unless the individual’s current environment changes, there is little likelihood that any “potential” can actually be realized. We have shown that the ARLT, by differentiating gainers from nongainers in a population classified as moderately to severely mentally retarded, leads to useful and beneficial information for educational purposes (consequential validity, Reschly, 1997). It facilitates the establishment of differentiated educational programs and the knowledge necessary to “push” the students who prove that they are able to benefit from a cognitive training (gainers). Yet, when they continue to receive the same schooling as the nongainers, their latent competencies will fade out instead of being developed. As Landesman Ramey et al. (1997) stated, “an environment that fails to demand increasingly mature and independent responses from a child may restrict development by not providing problems to solve and by not offering feedback about the child’s developmental accomplishments” (p. 64). Conversely, if unreasonable high demands are placed on students, they are likely to experience repeated failures, which, in turn, may decrease their willingness to engage in such activities. The deficits and difficulties of individuals in this population are well-known, and trying to be more demanding with regard to teaching for all students would be a loss of energy for teachers and could result in a loss of self-confidence and self-image for the weakest students.

For those students who show an undetermined status on the ARLT, opportunities for learning with more academic or cognitive content should be provided too. Indeed, some of them in our research profited from training, whereas others did not. The learning capacity of such individuals is difficult to estimate even with our test. It seems that the major component needed here is time. Some of these students appear to be slow gainers, with a potential that needs time to materialize. In this study, it was only after the ARLT administration and the one-month training that a number of them showed improved performance.

With regard to the nongainers, it seems that they, indeed, cannot surpass a concrete level of reasoning, as argued by Jensen (1969, 1970), and the usual educational setting seems to be appropriate for them. However, it is stressed that the training in the present study was relatively short. Further research is needed to determine whether the reasoning abilities of the nongainers can be enhanced with longer training and perhaps other kinds of intervention. In a follow-up study, Büchel (1999) showed that some of the nongainers could improve their reasoning abilities after a one-
year training, but that their short-term memory span remained an important limiting factor that could not be improved.

The ARLT can also help to overcome low expectations of teachers, educators, and psychologists with regard to individuals whose IQs are below 55. Teachers with low expectations make less rigorous demands and are more easily accepting of poor performance, which may result in a Pygmalion effect. They may fail to recognize that higher cognitive skills can be learned. Some studies have illustrated how the expectancies of teachers have changed after having been informed about their pupils’ results on dynamic tests (Hessels, 2000; Vye, Burns, Delclos, & Brandsford, 1987).

Another advantage of the ARLT lies in cognitive and aptitude by treatment research. In the few investigations that have been focused on this area, often no or few effects of cognitive training could be found with participants whose IQs were below 55. It is argued here, as Wanshura and Borkowski (1975) already outlined, that training effects can be underestimated when only the means are considered and, thus, individual variability is hidden. Results of this research show that some individuals with moderate or severe mental retardation can profit from cognitive training and some cannot. To test the effect of cognitive training for individuals with moderate or severe mental retardation, one must consider inter-individual variability, such that the scores of the nongainers do not mask the scores of the gainers. It would be wise to first distinguish the participants with regard to their learning capacities. It can be assumed that in this case, the program would be effective for some of the participants (gainers), but not for others (nongainers), which would prove the educability of some of these individuals. This is essential for interaction aptitude by treatment research, which is largely missing for this population.

Jensen (1992) distinguished two models with regard to educational decisions: The stability and the change model. The first, derived from the psychometric tradition, considers cognitive abilities as fixed. In this perspective, the educational implications are that the learning environment of individuals with low IQ should be changed (i.e., place them in less demanding settings). In the change model, also adopted by other authors (Campbell & Carlson, 1995; Carlson & Wiedl, 1992; Feuerstein, Hoffman, Jensen, & Rand, 1985; Haywood & Switzky, 1992), cognitive abilities are considered as modifiable. Given this perspective, educational settings should address existing abilities in order to change cognitive functioning. It is admitted that development is influenced by cognitive, metacognitive, and motivational variables. It is time that educational programs for individuals with moderate or severe mental retardation become more differentiated and more challenging for those who could benefit from training. As previous researchers have shown, individuals with various etiologies display different characteristics (e.g., a relative strength in visual modality compared to auditory modality, relative strength in simultaneous processes as compared to sequential processes). Educational interventions could be more effective when taking these into consideration (Hodapp & Dykens, 2001). These authors cited, for example, promising studies in which investigators found a significant effect of a visual instruction (the participants had to read the instructions) as opposed to auditory instruction (the examiner told the participants the instructions) for children with Down syndrome on vocabulary, grammar, and memory tasks.

Yet, few data are available concerning educational programs for students with moderate or severe mental retardation. Considerable work is still needed in order to develop useful programs, didactics, and material. In some cases, the Instrumental Enrichment Program of Feuerstein and colleagues (e.g., Feuerstein, Rand, Hoffman, & Miller, 1980) has successfully been used with this population (Lifshitz & Rand, 1999). However, these kinds of cognitive programs often lack significant guidelines for the cognitive remediation, a criticism noted by Lauchlan and Elliott (1997). It is hoped here that because the present study has clearly shown that some of the individuals with moderate or severe mental retardation can access a rather abstract level of reasoning, further research will be devoted to the devices of adapted curricula and technologies to bring about changes in cognitive functioning.

References


Assessing learning capacity

C. Hessels-Schlatter


Received 4/12/01, accepted 10/17/01.

This study was supported by a research grant (11–36345.92) awarded to F. P. Büchel by the Swiss National Scientific Research Fund. Requests for reprints should be sent to Christine Hessels-Schlatter, Faculty of Psychology and Educational Sciences, Bd du Pont-d’Arve, 40, 1205 Geneva, Switzerland. E-mail: Christine.Hessels@pse.unige.ch.