

# CAN MENTAL RETARDATION BE REVERSED?

JOHN M. THRONE and JOEL FARB

University of Kansas, Lawrence, Kansas 66045, U.S.A. and  
Lakemary Centre for Exceptional Children, Paola, Kansas 66071, U.S.A.

The question is: Can mental retardation be reversed? The answer appears to be yes: (1) if mental retardation is defined behaviourally; (2) if the behaviour in question is identified by performances on tasks like those found on standardised intelligence tests but not elicited under the neutral conditions of standardised testing; (3) if, on the contrary, such performances are deliberately and systematically trained to achieve non-retarded levels of proficiency; (4) if such gains hold over time; and (5) if such results generalise from the trained performances to untrained performances from the same or related categories.

## Intelligence Training

Given these five conditions, evidence for the reversibility of mental retardation is contained in a series of reports on research with which the authors have been involved since 1974. (Cf., Farb & Throne, 1974; Farb, Throne, Sailor, & Baer, 1974; Cottrell, 1975; Cottrell, Montague, Farb, & Throne, 1976; Farb, 1976; Farb, Cottrell, Montague, & Throne, 1977; Throne, Farb, Cottrell, & Montague, 1977.) In these studies mental retardation is attacked through intelligence training. To be retarded is to be relatively low in intelligence in comparison to the mean of the population. Therefore, raising the intelligence of a retarded individual toward the population mean is to reverse his or her retardation.

But intelligence is a multiple, not a unitary, concept. We have used as indices of its multiple components, performances from, and like those tapped by, the 12 subscales of the Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949; 1973), since all our subjects up to now have been children. Our choice of the WISC has been arbitrary but not capricious; it is probably the foremost standardised intelligence test for children in the world.

Note, we have said we use intelligent performances to indicate the multiple components of intelligence. We regard intelligent performance and intelligence as scientifically identical. Distinguishing between them cannot be done scientifically because their parameters are identified by the same, behavioural, data. We solve the problem of the relationship between them by equating intelligence, which is an abstraction, with intelligent performance, which is concrete. The practical upshot of this solution is to free us from hesitation over where our attack on mental retardation should begin: it is with sub-par intelligent performances, as inseparable from sub-par intelligence as energy is from mass times the speed of light squared.

Through sub-par intelligent performance, then, our attack on mental retardation is launched across the terrain of mentally retarded behaviour. By defining retardation behaviourally we satisfy the first condition for reversing it. We have already mentioned how we meet the second, third, fourth, and fifth conditions. We identify retarded behaviour by performances on tasks from and like those found on standardised intelligence tests but do not leave these performances to be elicited, if or howsoever they may be, under the conditions of standardised testing (second condition). We subject these performances to the strategies and tactics of operant conditioning (Skinner, 1953), the deliberate and systematic manipulation spoken of above, in hot pursuit of non-retarded performance levels (third condition). We aim not for the acquisition of non-retarded performances in the short-run but for their maintenance

over the long-run (fourth condition), as well as for generalisation of training to untrained performances to which the trained performances are categorically related (fifth condition).

## **Research into Increasing the Intelligence Levels of Mentally Retarded Children**

### **General Methodology**

Specific intelligent performances like those found on the WISC were selected for training.<sup>1</sup> In order to determine the extent to which improvement in performances produced through training generalised, performances related to them categorically served as probes. Probes, of which there may be several sets, are performance items which never receive training, but are systematically observed for evidence of generalisation. Generalisation is determined by the extent to which changes in probe performances are functionally related to changes produced through training of targeted performances (Baer, Peterson, & Sherman, 1967).

As all our studies were concerned with performances of individual children, single subject designs (Hall, 1971) were applied throughout them all. In particular, the multiple baseline design (Baer, Wolf, & Risley, 1968) was utilised. Use of the multiple baseline design allows not only for continuous monitoring of training effects but also allows for continuous monitoring of its generalised effects as evidenced by the probe measures.

Training procedures involved application of operant techniques (Skinner, 1953) such as shaping, chaining, and fading to the targeted intelligent performances under investigation. How these techniques were applied, and with what results, is summarised below.

### **Specific Methodology and Results<sup>2</sup>**

Performances representing three intelligence components — that is, three of its categories as measured by WISC subscales — were targeted for training. They were memory, vocabulary, and spatial relations.

#### **Memory**

Improving the mnemonic(memory) component of intelligence was the goal of our initial studies. The first study (Farb & Throne, 1974; Farb, 1976) was aimed at the mnemonic performances of a Down's Syndrome girl. The subject, a student at Lakemary Centre for Exceptional Children, was 6 years and 7 months of age at the onset of the study. Her I.Q.'s on standardised intelligence tests consistently clustered around 50.

The specific set of mnemonic performances selected for training was digit-span. A pool of digit-span items similar to those on the WISC was developed. These items varied in length from three to five digits. They were divided into item classes on the basis of length. Hence, three-digit items formed the basis of the three-item class, four-digit items the four-item class, five-digit items the five-item class. Items to be trained were drawn from the pool for each item class.

Each item class also contained four sets of probes. The first set consisted of digit-span items similar to those toward which training was directed, and included the digit-span items from the WISC. The second set of probes consisted of grammatical sentences corresponding in length (number of syllables) to the number

<sup>1</sup>Items from the WISC itself were reserved as probes. (See below.)

<sup>2</sup>Space limitations preclude inclusion of 29 figures and tables in this report. They are available upon request from the first author.

of digits in each of the digit-span item classes. The third set of probes consisted of nongrammatical sentences corresponding in length (like the grammatical-sentence probes) to the digit-span item classes. The fourth set of probes consisted of match-to-sample items requiring that the subject point to the letters on a card matching those on a sample card which had been removed after having been presented to the subject for viewing. The number of letters to be matched corresponded to the various digit-span item classes (i.e., 3-5). It should be remembered that the probe performances never received training, but served only as generalisation measures.

Baseline measures were recorded across all performances. Following the recording of baseline, training was instituted for one item class at a time in keeping with a multiple baseline design across item classes (Hall, 1971). The training procedure consisted of modeling digit-span items with reinforcement contingent upon a correct imitation of the item within 10 seconds following presentation of the item. Reinforcement<sup>3</sup> consisted of a combination of praise and an edible. If the subject's response to a digit-span training item was incorrect, the item would be repeatedly presented with reinforcement contingent upon increasingly correct responses until a correct response to the item was emitted. Responses to probe items produced no programmed consequences but were simply recorded as either correct or incorrect. The percentage of correct responses per session was recorded for all performances, training and probe.

In the first study, two forms of the training procedure were investigated. In the first form, all items, both training and probe, were the same from day to day. In the second form, all items were presented in continually varying order and differed from day to day. The first form of the training procedure resulted in the percentage of correct responses to the digit-span training items increasing to 100% in each of the item classes trained (3-5). However, generalisation as evidenced by the probe measure was minimal. The second form of the procedure resulted in the percentage of correct responses to the digit-span training items increasing to levels substantially above baseline with corresponding increases in performances on the probes. Hence, the second form of the training procedure was successful in producing not only improved mnemonic performances but generalised improvements. At the termination of training, the subject's mnemonic performances were at a level indicated by the WISC to be normal for a child of her chronological age. A follow-up showed all improved performances to be maintained.

An additional study was conducted in which the second form of the training procedure was applied with two other children (Farb, Throne, Sailor, and Baer, 1974). The first of these two subjects was 10 years of age at the onset of the study and was diagnosed as culturally deprived as well as mentally retarded. His I.Q.'s clustered around 60. The second subject was 11 years and 9 months of age. He was a Down's Syndrome child whose I.Q.'s clustered around 45. For the first subject, the training procedure was applied to item classes five through seven. For the second subject, item classes three through five were trained. Along with the digit-span training items, the same set of probes utilised in the first study was also examined for evidence of generalisation.

Again, the training procedure resulted in improved performance levels across all measures training and probe. For the first subject, training was terminated with the seven-item class — normal for a child of his chronological age. For the second subject, training terminated with the five-item class — slightly below normal for a child of his chronological age. A follow-up probe session conducted seven months after the termination of training showed all effects to be maintained. The results of this second study thus duplicated the results obtained in the first study.

<sup>3</sup>Strictly speaking, one should not speak of reinforcement until a stimulus following a response is shown to raise the frequency, percentage, or rate of occurrence of responses in the same response class. The stimulus may then be designated a reinforcer.

## Vocabulary

While the studies involved with training mnemonic performance were in progress at Lakemary, two colleagues at the University of Arkansas at Little Rock, Ann Cottrell and James Montague, conducted a study aimed at improving vocabulary performance (Cottrell, 1975; Cottrell, Montague, Farb, & Throne, 1976). The subject in this study was a Down's Syndrome girl who was 6 years and 6 months of age at the onset of the study. Her I.Q. on the WISC was 52, and she scored an I.Q. of 42 on the Peabody Picture Vocabulary Test.

The first eight words of the WISC vocabulary subscale were used as the basis for eight semantic classes. For example, the WISC word **bicycle** was treated as a member of the semantic class **vehicles**; **knife** as a member of the class **cutting instruments**; **hat** as a member of the class **clothing**; etc. In each semantic class, three words were developed to serve as training items. The WISC word in each semantic class served as a probe and never received training. In other words, the semantic class **vehicles** (as an example) involved the words **bus**, **boat**, and **plane** as training words while **bicycle**, the WISC word, served as a probe for measuring generalisation.

All of the words in each semantic class were treated in terms of both identification and definition (e.g., "Show me the boat," and "What is a boat?" respectively). Training correct identification was conducted together with training correct definition. Trials consisted of asking the subject for either an identification or a definition response in the manner noted above. The same procedure was conducted for the WISC probe word. The percentage of correct responses was recorded for the identification and definition trials.

Baselines were recorded across all words in all semantic classes. Following the recording of baselines, training proceeded progressively from one word to another in multiple baseline fashion within the first semantic class. Then training on words in the next semantic class proceeded so that the experimental design involved a multiple baseline across words within the larger context of a multiple baseline across semantic classes (Hall, 1971). The probe words in each semantic class were observed for evidence of generalised effects as training proceeded for each semantic class. Training proceeded on a given semantic class until the percentage of correct responses to the training words had increased to criterion (i.e., three successive sessions at or about 80%) along with a corresponding increase in correct responses to the probe word.

An identification trial would be conducted in the following manner. If **boat** was the word being presented, then pictures representing all four words in the semantic class (**bus**, **boat**, **plane**, and **bicycle**) were placed on a table in front of the subject. The subject was then instructed, "Show me the **boat**." A correct response consisted of the subject pointing to the picture of the boat within ten seconds following the instruction. A definition trial would be conducted by asking the subject (for instance), "What is a boat?" Correct responses to the definition trials were scored in accordance with the criteria listed for the WISC vocabulary subscale. Probe words were presented in the same manner as the training words; again, responses to these items, whether correct or incorrect, produced no programmed consequences.

The training procedure involved reinforcement contingent upon correct responses to the training words and/or successive approximations of a correct response. The reinforcer consisted of a combination of praise and an edible. Incorrect responses resulted in the trial being presented again with the correct response being modeled for the subject by the trainer. Successive approximations of the correct response were reinforced over repeated trials until a correct response was emitted and reinforced. Responses to probe words were never repeated or reinforced but simply recorded as either correct or incorrect.

The procedure resulted in rapid vocabulary gains that generalised within each semantic class as it was applied. It was often found that training on only the first two words within a given semantic class produced generalisation of correct responses to the remaining training word and the probe word in the class.

The results of the study also demonstrated that the subject could frequently identify the pictorial referent of a word while being unable to define the word prior to training. At no time was correct identification at a lower level than correct definition. While neither identification nor definition responses were trained for the probe words, criterion level increases in correct responses occurred on both measures for the probe words in each semantic class. Similar improvements in vocabulary performance were obtained when the training procedure was applied in behalf of two other children (Farb, Cottrell, Montague, & Throne, 1977; Throne, Farb, Cottrell, & Montague, 1977), thus duplicating the results observed for the first subject.

A follow-up probe session was conducted with all three subjects 90 days after the termination of training. The results of the follow-up showed all gains, for all subjects, to have been maintained over the interval during which no training sessions had been conducted.

### Spatial Relations

The final study in this series to date was aimed at the improvement of the spatial relations component of intelligence. The specific objective of the study was the improvement of block-design performance (Farb & Throne, 1977; Cottrell, Montague, & Throne, 1977; Throne, Farb, Cottrell, & Montague, 1977). The study involved two retarded girls at Lakemary. The first subject was 10 years and 8 months of age at the onset of the study. Her I.Q.'s clustered around 70. The second subject was 14 years and 5 months of age. Her I.Q.'s clustered around 50. In both cases, brain damage was indicated.

The study involved one set of block-design training items and four sets of probe measures. The block-design training items consisted of block-design tasks similar to but different from the block-design items on the WISC. The first set of probes consisted of block-design items similar to but different from the training items and included the block-design items from the WISC. The block-design training items and the block-design probes were presented on index cards showing a given block-design. In constructing the block-designs, the subjects used the blocks from the WISC block-design subscale. A correct response consisted of the subject constructing the pictured block-design within 65 seconds or within the maximum time criterion set by the WISC in the event of a probe taken directly from the WISC block-design subscale.

The second set of probes consisted of two-dimensional block-constructions. The third set of probes consisted of three-dimensional block-constructions that were otherwise exactly like the two-dimensional probes. The number of blocks involved in the block-construction probes was determined for each subject on the basis of his respective baseline levels. The block-construction probes involved seven blocks for the first subject and five blocks for the second subject. The block-construction probes were presented on the index cards. The index cards showed either a two-dimensional block-construction or a three dimensional block-construction. A correct response to either block-construction probe involved the subject duplicating the pictured construction within 65 seconds using the blocks from the Stanford-Binet Intelligence Scale, L-M (Terman & Merrill, 1964).

The fourth set of probes consisted of "figural-reasoning" match-to-sample tasks. In presenting a "figural-reasoning" probe, the subject was shown two cards. On the first card were three abstract designs. On the second card were three other abstract figures. A correct response consisted of the subject pointing to the figure on the second card which belonged with the figures shown on the first card.

Baseline measures were recorded across all items for both subjects. Following the recording of baselines, training was instituted with the first subject while the second subject continued in the baseline condition in accordance with the format of a multiple baseline design across subjects (Hall, 1971).

The training procedure involved reinforcement, in the form of praise, contingent upon correct responses and/or successive approximations of correct responses to the block-design training items. When a block-design training item was presented, each correct placement of an individual block produced reinforcement. If the entire block-design item was completed correctly within the time criterion, then reinforcement was again provided and the response was recorded as correct. If an individual block was placed incorrectly in the process of performing a block-design item within the time criterion, no reinforcement was forthcoming. If the error was not corrected by the end of the time criterion (in which case no reinforcement would be provided), the response was recorded as incorrect and the trial was repeated. When a trial was repeated, an incorrect block placement resulted in the trainer correcting the error by placing the block in the correct position. The trainer then replaced the block in its original incorrect position. Reinforcement was contingent upon the subject then performing the correction which had been modeled by the trainer. On any trial during which such training occurred, the response for the trial was recorded as incorrect. This procedure was repeated until the subject performed a complete block-design without error. However, if a complete correct response to an item did occur on the fourth presentation of the item, another training item was presented and the previous item was presented again at a later time. Responses to probes, whether correct or incorrect, produced no programmed consequences. All items, both training and probe, varied from session to session.

The introduction of the training procedure with the first subject resulted in rapid improvements in block-design performances with the percentage of correct responses to all items, training and probe, increasing to levels substantially above baseline. Seven days after the beginning of training for the first subject, training was initiated for the second subject, with the results showing similar rapid improvement across all items, training and probe. As in the memory and vocabulary categories, improvements of trained items in the spatial relations category were accompanied by improvements of untrained items in the category as well, indicating generalisation.

### **Conclusion**

To conclude, research conducted since 1974 supports the affirmative answer to the question regarding the reversibility of mental retardation raised in the title of this article. This follows only if the five assumptions under which this research has been carried out are accepted. To date, our research has been concerned with the investigation and demonstration of systematic procedures for the improvement of intelligent performances corresponding to the WISC categories of memory, vocabulary, and spatial relations. The most significant information yielded by these studies is not only that training of targeted intelligent performances, using operant conditioning, results in improvement of these performances which hold over time, but that performances categorically related to them that remain untrained also improve. In other words, intelligence training generalises. We believe these findings could have profound implications for the mentally retarded in the future.

Research now planned is aimed toward investigating and demonstrating procedures for increasing performance levels in each of the nine remaining intelligence categories contained in the WISC. Once we have arrived at this goal, all twelve sets of training procedures, corresponding to the twelve WISC categories, will be presented to individual retarded children. We have proposed a research design (Throne & Farb, 1975; Throne & Farb, 1977) that should enable us to determine the functional effects of all these training procedures within and across all intelligence

categories corresponding to the 12 WISC subscales. In addition to intelligence measures, probes will include academic achievements, social adaptation, and other measures closely related to several if not all components of intelligence (probably) though not usually regarded to fall within its parameters per se. To the extent that this research "package" is successful, we believe we may be on the way to convincing even those inclined to be most sceptical, not only that mental retardation can be reversed, but that steps should be taken to ensure that it is reversed (Throne, 1975; 1977).

### Summary

Five conditions are listed under which the possibility of reversing mental retardation is affirmed. Recent research is summarised suggesting that reversing mental retardation may be successfully undertaken.

### References

- Baer, D. M., Peterson, R. F., & Sherman, J. A. The development of imitation by reinforcing behavioural similarity to a model. *Journal of the Experimental Analysis of Behaviour*, 1967, 10(5), 404-416.
- Baer, D. M., Wolf, M. M., & Risley, R. R. Some current dimensions of applied behaviour analysis. *Journal of Applied Behaviour Analysis*, 1968, 1(1), 91-97.
- Cottrell, A. W. An operant procedure for improving the vocabulary performances of a mentally retarded child. Unpublished master's thesis, University of Arkansas, Little Rock, 1975.
- Cottrell, A. W., Montague, J. C., Farb, J., & Throne, J. M. An operant procedure for improving vocabulary definition performances in developmentally delayed children. Paper presented to the annual meeting of the Arkansas Speech and Hearing Association, Jonesboro, Arkansas, March, 1976.
- Farb, J. Operant investigative training of generalised mnemonic performances of a Down's Syndrome child. Unpublished master's thesis, University of Kansas, Lawrence, 1976.
- Farb, J., Cottrell, A. W., Montague, J. C., & Throne, J. M. Update on research into increasing intelligence levels of the mentally retarded. In P. Mittler (Ed.), *Research to practice in mental retardation: Proceedings of the International Association for the Scientific Study of Mental Deficiency*. Volume 2. Baltimore: University Park Press, 1977, 129, 129-136.
- Farb, J. & Throne, J. M. Operant investigative training of mnemonic performances of a Down's Syndrome child: A preliminary report. Paper presented to the joint meeting of the American and Canadian Associations on Mental Deficiency, Toronto, June, 1974.
- Farb, J. & Throne, J. M. Generative intelligence: Improving generalised block-design performance of mentally retarded children. Paper presented to the annual meeting of Region V of the American Association of Mental Deficiency, San Antonio, October, 1977.
- Farb, J., Throne, J. M., Sailor, W., & Baer, D. M. Operant investigative training of mnemonic performances of mentally retarded children. Paper presented to the annual meeting of Region V of the American Association on Mental Deficiency, New Orleans, October, 1974.
- Hall, R. V. *Managing behaviour, Part 1. Behaviour modification: The measurement of behaviour*. Lawrence, Kansas: H & H Enterprises, Inc., 1971.
- Skinner, B. F. *Science and human behaviour*. New York: MacMillan, 1953.
- Terman, L. M. & Merrill, M. A. *Stanford-Binet intelligence scale*. Boston: Houghton-Mifflin, 1960.
- Throne, J. M. Raising intelligence levels of the mentally retarded: An overlooked educo-legal implication. *Journal of Education*, 1975, 157(1), 43-53.
- Throne, J. M. Increasing intelligence through training: A future role for school and residential facilities for the retarded. *Mental Retardation*, 1977, 15(6), 25-27.
- Throne, J. M. & Farb, J. An operant alternative to multivariate statistics: A proposal. *Educational Technology*, 1975, 15(9), 27-32.
- Throne, J. M. & Farb, J. Operant multivariate analysis. In J. S. Martois (Ed.), *Proceedings of the Fourth Annual Conference and Exhibition on Measurement and Evaluation: Office of the Los Angeles County Superintendent of Schools and the Los Angeles County School Board*. Downey, California: OLACSS-LACBE, 1977, 149-155.
- Throne, J. M., Farb, J., Cottrell, A. W., & Montague, J. C. Repeat of extended update on research into increasing intelligence levels of mentally retarded children. In J. S. Martois (Ed.), *Proceedings of the Fourth Annual Conference and Exhibition on Measurement and Evaluation: Office of the Los Angeles County Superintendent of Schools and the Los Angeles County School Board*. Downey, California: OLACSS-LACBE, 1977, 161-198.
- Wechsler, D. Wechsler intelligence scale for children. New York: The Psychological Corp., 1949.
- Wechsler, D. Wechsler intelligence scale for children—Revised. New York: The Psychological Corp., 1973.