Effects of Iron Treatment on Cognitive Performance and Working Memory in Non-anaemic, Iron-deficient Girls

Anthony Lambert, Karrin Knaggs
Research Centre for Cognitive Neuroscience, University of Auckland

Robert Scragg, Patricia Metcalf, David Schaaf
School of Medicine, University of Auckland.

A randomised, double-blind intervention study is described in which 116 adolescent girls with iron deficiency in the absence of anaemia were randomly allocated to iron treatment or a placebo control group. Both cognitive performance and haematological indices of iron status were assessed before and after an eight week treatment period. The four cognitive tests comprised: *Hopkins Verbal Learning Test* (HVLT), *Stroop Task*, *Visual Search*, and *Reading Span Task*. Between pre- and post-treatment testing sessions participants in the Iron group showed significant improvement in immediate recall of recently heard words from the Hopkins word lists, while those in the Placebo group showed no improvement. In addition, at post-treatment significant relationships were found between haemoglobin levels and recall of recent words and between serum ferritin levels and reading span. These results suggest: (a) that iron deficiency may impair the performance of everyday activities that involve verbal working memory, and (b) that improvement in iron status may bring about an improvement in verbal working memory performance.

Iron deficiency is an extremely common nutritional problem. According to Scrimshaw (1991) 10-20% of women of child-bearing age in the U.S., Europe and Japan are anaemic. In most developing nations prevalence rates are even higher. The aim of the study described below was to address the question of whether iron deficiency affects the cognitive performance of female school age adolescents. The potential importance of this issue in a New Zealand context is highlighted by results recently reported by Schaaf, Scragg, Metcalf, Grant and Buchanan (2000). These authors carried out a survey at eight Auckland high schools, and found relatively high prevalence rates of iron deficiency in female students, especially among Maori (25.6%), Pacific Island (20.9%) and Asian (15.9%) participants. Iron deficiency was somewhat less common in European female high school students (8.3%). Although Schaaf et al. (2000) were unable to identify specific dietary or other factors responsible for this ethnic variation, it is clear from their results that iron deficiency is very common among female high school students in New Zealand.

It is important to remember that though they are related, iron deficiency and anaemia are distinct conditions. Iron deficiency refers to a general depletion of iron stores in the body. Operationally, iron deficiency can be defined in terms of the concentration of ferritin in blood serum. According to the World Health Organisation (1972) concentrations less than 12mg/litre may be considered diagnostic of iron deficiency. Anaemia is a more serious condition which arises from prolonged iron deficiency and is manifested by a depletion of blood haemoglobin, which is present in red blood cells and is necessary for oxygen transport. According to the World Health Organisation (1972) haemoglobin levels less than 120g/litre may be considered diagnostic of anaemia.

It is well known that anaemia has widespread bodily consequences, including reduced aerobic capacity, increased fatigue and feelings of lassitude, and impaired cognitive performance (see Scrimshaw, 1991). However, it is not clear whether the milder condition of iron deficiency in the absence of anaemia has important consequences for cognitive performance. This issue was addressed in a study by Bruner, Joffe, Duggan, Casella and Brandt (1996), which examined effects of iron supplementation on the cognitive performance of non-anaemic iron-deficient adolescent girls in Maryland, USA. This important study found that compared to a placebo group, participants who received iron tablets (1300mg of ferrous sulphate - equivalent to 260mg of elemental iron - per day for 8 weeks) improved their immediate recall performance on the *Hopkins Verbal Learning Test* (HVLT). The study showed that simple treatment of iron deficiency with a dietary supplement was associated with an improvement in memory performance. However, these results also prompt a number of further questions. Although Bruner et al. found that iron treatment was associated with improved ability to immediately recall lists of 12 words, there was no association with delayed recall or recognition. It is widely thought that immediate recall performance of the kind studied by Bruner et al. reflects the operation of two distinct kinds of memory process. It is thought that participants draw upon a limited capacity short term or working memory system when recalling items from the end of the list. In contrast, recall of earlier items is thought to be achieved primarily by retrieving them from long term memory. Accordingly, it is unclear whether the significant effect of iron treatment reported by Bruner et al. reflects an improvement in working memory performance, or an improvement in storage or retrieval from long term memory. The absence of effects on delayed recall or recognition might be seen as tentative evidence in favour of a working memory interpretation of their finding. The aim of the present study was to explore further the relationship between iron status and cognitive performance, especially with respect to working memory.

Our study resembled that of Bruner et al. (1996) in following a double-blind intervention design. Female adolescents suffering from iron deficiency, but not anaemia, were randomly assigned to one of two treatment groups, iron or placebo. Both cognitive performance and haematological measures of iron status (blood haemoglobin and serum ferritin concentrations) were assessed before and after an eight week treatment period. During the eight week treatment period participants in the Iron group received dietary iron supplement tablets on a daily basis, while participants in the Placebo group received placebo tablets that were carefully matched for visual appearance. Following treatment it was expected that serum ferritin levels would increase for participants in the Iron group, but not for those in the Placebo group. However, a similar prediction was not advanced with respect to haemoglobin, levels since it is known that haemoglobin levels respond to iron treatment with a much slower time course than ferritin levels. Accordingly, it was anticipated that a period of eight weeks from the commencement of treatment may be insufficient to observe a significant increase in blood haemoglobin.

The hypothesis that the cognitive performance of our iron-deficient participants would be sensitive to iron treatment was tested in several ways. Firstly, this hypothesis predicted an improvement in the performance of the Iron group relative to the Placebo group across the two testing sessions. This primary prediction was tested with a split-plot analysis of variance, comparing the performance of the two participant groups across the two testing sessions. A secondary prediction was that the extent of this improvement may be related to the extent of improvement in iron status, as assessed by ferritin and haemoglobin levels. In a sense the multiple regression (MR) analyses used to test this prediction provided a more powerful test of the hypothesis than the simpler binary comparisons between treatment and placebo groups provided by analysis of variance. The reason for this is that within the treatment group it is very likely that for some participants iron levels may improve only slightly, or perhaps fail to improve at all across the treatment period. This could result from poor compliance (i.e. failing to take the tablets) or poor absorption due to other dietary or health factors. Conversely, within the placebo group it is quite possible that iron levels may improve for some participants. Information provided to participants during the recruitment stage of the study would of course raise awareness of the nutritional importance of iron and may lead to an improvement in dietary iron intake. Multiple regression analysis was valuable in this context in capturing and testing for effects of fine grained changes in iron status. This fine grained information is not captured in the binary comparison of treatment and control groups provided by the primary analysis.

A further perspective here is that the analyses of variance and MR analyses can be viewed as providing answers to closely related, but perhaps distinct research questions. The binary comparisons provided by the former can be viewed as answering the essentially practical question that is of interest to clinicians, nutritionists and educators: does treatment of iron-deficient girls with dietary supplements produce measurable improvement in cognitive performance relative to a control group? The MR analyses address a slightly broader issue: is cognitive performance related to iron levels and to changes in iron status (irrespective of whether these are brought about by treatment or by other factors such as dietary improvements for some individuals within the Placebo Group)?

The more specific hypothesis that working memory may be related to iron status was tested in two ways. Firstly, free recall performance on the HVLT was evaluated separately for the first six items and the last six items in each word list. These will be referred to below as early items and recent items respectively. The working memory hypothesis predicts that iron treatment will be related to recall of recent words, but not to recall of earlier words. In addition, participants performed a test specifically designed to assess working memory: the reading span test of Daneman and Carpenter (1980; see also Masson & Miller 1983). An important reason for including this test in our battery was that there is evidence to suggest that performance on this test is strongly related to reading comprehension (Daneman & Carpenter, 1980). Clearly, reading comprehension is a core task necessary for academic achievement among our participant population – school age girls. It is likely therefore, that poor performance on Daneman and Carpenter's reading span test will be associated with academic under-achievement at school.

A surprising feature of the findings of Bruner et al. was that although iron treatment was associated with improved immediate recall, there was no relationship with any of the three measures of attention that were employed (*Brief Test of Attention, Visual Search and Attention Test, Symbol Digit Modalities Test*). This was surprising since an association between iron status and attention had been implicated in earlier studies. Pollitt, Soemantri, Yunis & Scrimshaw (1985) found that anaemic 8-11 year old children improved their performance in terms of both speed and accuracy on a visual attention task (*Matching Familiar Figures Test*) following iron treatment. In addition, Oski Honig, Helu and Howanitz (1983) found that following iron treatment non-anaemic iron-deficient infants (9-26months) showed significant improvement on the *Bayley Scale of Infant Development*, which assesses attention, reactivity and motor development. Bruner et al. speculated that lack of sensitivity in some of their measures of attention, together with ceiling effects may have obscured a relationship between iron treatment and attentional processes. They conclude that further research exploring this issue is required. Accordingly, the present study included two measures of attention: the well known *Stroop* colour word test, and a visual search task. In the *Stroop* colour word test participants named the colour of letter strings printed on a sheet of paper. In the Interference condition these letter strings were colour names that were discordant with the colour in which they were printed (e.g. the word RED, printed in green ink). In the Control condition the letter strings were rows of X's (e.g. XXX printed in green ink). The *Stroop* test is widely regarded as a test of executive attention, because by comparing the two conditions one can assesses the ability of participants to suppress the competing colour name response (see Posner, 1994). The visual search task involved measuring the speed and accuracy of searching for a target object among distractors. It was modelled on the conjunction search task of Treisman & Gelade (1980; see also Treisman, 1988). Participants searched for a red X, amongst distractors that included both red letters, and green X's.

Method

Participants

A total of 870 female students attending a high school in Auckland, New Zealand, volunteered for the screening phase of the study and provided venous blood samples that were analysed with respect to serum ferritin and blood haemoglobin. The blood samples of 666 participants (77%) were normal; 135 (16%) were iron-deficient without being anaemic (serum ferritin < 12 mg/l, haemoglobin > 120 g/l); 52 participants (6%) had iron deficiency with anaemia (ferritin < 12 mg/l, haemoglobin < 120 g/l); 17 (2%) had anaemia in the absence of iron deficiency. Of the 135 girls with non-anaemic iron deficiency, 121 volunteered to participate in the trial. Five dropped out in the course of the trial, leaving 116 participants who completed the study. The age of these participants ranged from 12.5yrs to 17.9yrs, with a mean of 15.2yrs. There were 57 participants in the Iron group and 59 participants in the Placebo group. Both the researchers and participants were blind to the composition of these groups until completion of the study.

Procedure

Participants completed the four experimental tasks in the following order: *Hopkins Verbal Learning Test* (HVLT), *Stroop Task, Visual Search, Reading Span Task.*

Hopkins Verbal Learning Test: This was administered according to the protocol described by Brandt (1991). That is, participants were asked to listen carefully to a list of 12 words, and upon completion of the list were to recall as many words as possible in any order. The words were read out at a rate of approximately one every two seconds. Correct responses were recorded by the experimenter. This procedure was repeated twice more. Approximately twenty minutes later (after completion of the *Stroop* and *Visual Search* tasks), delayed recall and recognition were assessed. Participants were asked to recall as many words as possible from the list of twelve read out earlier. In the recognition test, participants were told that a list of words would be read out, and they were to say 'yes' if the word was from the original word list, and 'no' if it was not. Participants were presented with twenty four words, twelve of which were from the original list. Of the twelve words not present in the original list, six belonged to the same category as list words, and six belonged to unrelated word categories. Participants received different parallel forms of the HVLT at pre-treatment and post-treatment testing sessions.

Stroop Task: Participants were asked to name aloud the ink colour of a series of letter strings printed on laminated cards measuring 297 x 210mm (A4). They were asked to do this as rapidly and accurately as possible. Participants began with two practice cards. The first (control card) contained 30 strings of X's printed in red, green and blue. The second (interference card) contained 30 colour words (RED, GREEN, BLUE) printed in an ink colour (red, green or blue) that did not correspond to the colour word. The number of characters in the control letter strings and the interference colour words was matched. Participants then performed the naming task with two interference and two control cards, each of which contained 45 letter strings. Half the participants completed these in the following order: Control Interference Interference Control; the other half were presented with cards in this order: Interference Control Control Interference. Naming time was recorded using a digital stop-watch, and the total number of naming errors were recorded.

Visual Search task: This was administered using an IBM compatible laptop computer. Participants were seated at a comfortable distance from the display screen and were informed that this was a test of perceptual speed. Participants were asked to search the display for the presence of a target object: a red X. The number of items presented on the screen (1, 2, 4, 8 or 16) varied randomly from trial to trial. Distractor letters comprised red Ts and green Xs. On 50% of trials the target was present, on 50% it was absent. Participants were instructed to press the 'n' key as quickly as possible if the target was present on the display, and the 'b' key if it was absent. When an error was made the word 'ERROR' was presented in the centre of the display for 1s. Participants pressed the spacebar in order to proceed to the next trial. Participants initially performed 10 practice trials, followed by four blocks of 40 experimental trials. The visual display and timing software for the experiment was written in *Turbo Pascal v.7.0.* Search time (accurate to the nearest millisecond) was calculated for present and absent decisions with each display size. The total number of errors was also recorded.

Reading Span Task: Participants were required to read out loud a set of unrelated sentences, and hold the last word of each sentence in memory. Initially, participants were given two practice trials, each of which involved reading two sentences, and then recalling the final words of each sentence. Representative examples of the sentences are: "*A telephone call may travel by wire or radio, and is sometimes bounced off satellites*", "*In the Middle Ages kings built castles to defend their land from enemies*". Following the practice trials, participants were presented with sentence sets of increasing difficulty level. Three sets of sentences were presented at each difficulty level. There were five levels of difficulty that contained two, three, four, five and six sentences respectively. The test began at the two sentence level. Each participant was given three trials of two sentences. If the final words of these sentences were recalled correctly, the participant progressed to the three sentence level. Participants were presented with increasingly longer sets of sentences until they failed to recall the final words in all three trials at a particular level. The test ended at this point. The test was scored in the following way. If a participant failed to recall the final words in all three trials at the five sentence level, but recalled correctly on one out of the three trials at the previous level, then a reading span of four would be recorded. If however, there was correct recall on two or three of the four sentence trials, a reading span of 4.5 would be recorded.

After completion of these tests at the first (pre-treatment) session, participants were given a bottle containing seven tablets. They were asked to take one tablet per day, preferably before breakfast. For participants in the Iron group, each tablet of Ferrogradumet (Abbott) contained a daily dose equivalent to 105mg of elemental iron. If a tablet was missed, two were to be taken on the next day. For the following eight weeks, participants would visit the school nurse once a week, to collect their tablets. The number of tablets taken during each week was recorded, as were any side effects that had occurred. At completion of the final bottle of tablets, the second (post-treatment) testing session was administered, and a second venous blood sample was taken, to determine ferritin and haemoglobin levels.

At a debriefing meeting following completion of all the cognitive tests participants were informed of initial results from the study. Each participant received a letter informing them as to what treatment group they were in, and of their post-treatment ferritin and haemoglobin levels. Participants in the placebo group received an eight week supply of iron tablets after the trial, to ensure that they were not disadvantaged relative to the Iron group.

Results

Effects of treatment on iron status

Effects of iron supplementation on the two haematological measures of iron status are shown in Table 1. Two split-plot analyses of variance were used to evaluate effects of the treatment on serum ferritin and haemoglobin levels respectively. Each of these analyses had one repeated measures factor (Testing Session: Pre-treatment vs. Post-treatment) and one independent groups factor (Treatment Group: Iron vs. Placebo). With respect to ferritin levels, effects of Testing Session ($F(1,114)=100.16$, $p<.001$), Treatment Group ($F(1,114)=58.81$, $p<.001$), and the interaction between these factors ($F(1,114)=53.07$, $p<.001$) were all highly significant. Further analysis showed that ferritin levels increased for both the Placebo Group ($t(58)=3.73$, $p<.001$), and the Iron Group ($t(56)=9.16$, $p<.001$). The significant interaction term indicates that the extent of this increase is considerably larger for the iron group.

Table 1. Effects of iron treatment on ferritin and haemoglobin levels

	FERRITIN LEVELS					
	Iron Group			**Placebo Group**		
	Pre-treatment	**Post-treatment**	**Change**	**Pre-treatment**	**Post-treatment**	**Change**
Mean	8.25	25.28	17.03	7.81	10.07	2.26
S.D.	2.3	14.3	14.0	2.2	5.5	4.6
Range	9	86	85	8	26	25
(Min .. Max)	2 .. 11	9..95	2..87	3 ..11	3..29	-5 .. 20
	HAEMOGLOBIN LEVELS					
	Iron Group			**Placebo Group**		
	Pre-treatment	**Post-treatment**	**Change**	**Pre-treatment**	**Post-treatment**	**Change**
Mean	131.3	132.5	1.2	130.1	126.5	-3.6
S.D.	6.8	7.7	6.4	7.0	8.9	6.8
Range	29	40	27	32	38	29
(Max...Min)	120 .. 149	115 .. 155	-11 .. 16	120 ..152	110 .. 148	-16 .. 13

With respect to haemoglobin levels, the effect of Testing Session did not attain significance (F(1,114)=3.86, p<.10). However, the main effect of Treatment Group (F(1,114)=15.34, p<.001) and the interaction between Treatment Group and Testing Session (F(1,114)=7.99, p<.01) were both significant (see Table 1). Further analysis showed that for the Placebo Group there was a significant decrease in haemoglobin levels between pre- and post-treatment, t(58)=4.09, p<.001. For the Iron Group there was no change in haemoglobin between pre- and post-treatment, t(56)=1.41, n.s.

Effects of iron treatment on cognitive performance

Effects of iron treatment were evaluated by examining performance data separately for each of our four cognitive tasks: *HVLT, Reading Span, Stroop Task*, and *Visual Search*. Three forms of analysis were employed. Our primary analyses for treatment effects were performed via split-plot analyses of variance which compared the performance of Iron and Placebo groups at pre- and post-treatment. The hypothesis that between pre-and post-treatment the performance of the Iron group would improve relative to the Placebo group was tested by the interaction term (Treatment Group x Testing Session). Since we did not consider the possibility that iron-deficient participants would perform more poorly as a result of iron supplementation, directional tests were used in evaluating this hypothesis.

Secondly, for each cognitive task multiple regression analyses were performed (using SPSS 9.0); each of these tested for relationships between performance and changes in iron status. In each case the dependent variable was performance at post-test; the predictor variables were performance at pre-test, age, change in ferritin level, and change in haemoglobin level. That is, each analysis tested for relationships of ferritin change and haemoglobin change with post-test performance, while controlling for pre-test performance and participant age. Since multiple regression analysis is notoriously susceptible to unwanted distortion due to the presence of outliers, the data were screened for outliers and observations with undue influence on the regression calculations. In doing this, we followed the three-fold recommendations of Judd and McClelland (1989). The following values were calculated with respect to each set of observations: Centred Leverage Values, Studentised Deleted Residuals, and Cook's D. If necessary, appropriate steps were taken to ensure that: (1) None of the 116 leverage values exceeded 0.2; (2) 95% of the studentised deleted residuals fell within the range +2 ... -2; (3) None of the Cook's D values exceeded 2.

Thirdly, in cases where multiple regression indicated significant relationships between haemoglobin or ferritin change, we tested for simple bivariate correlations between post-treatment performance and absolute levels of serum ferritin and haemoglobin following treatment.

Hopkins Verbal Learning Test: Immediate free recall performance on the Hopkins test is shown in Figure 1. The upper panel shows immediate free recall of words from the first part of the list (Serial Positions 1-6); the lower panel shows recall of words from the second half of the list (Serial Positions 7-12). An initial analysis of variance included the factor of Serial Position (First Half vs. Second Half) in addition to Treatment Group and Testing Session. The working memory hypothesis predicts that relative improvements in the performance of the Iron Group will be restricted primarily to recall of items from the latter half of the list. The hypothesis therefore predicted a three-way interaction between Treatment Group, Testing Session and Serial Position. This interaction was marginally significant on a directional test, F(1,114) = 1.83, p<.09. Further analysis showed that the interaction between treatment group and testing session approached significance for recall of recent words (F(1,114)= 1.76, p<.10 (directional). but not for recall of early items (F<1). Consistent with the working memory hypothesis, participants in the Iron Group showed significant improvement in recall of words from the second half of the list between pre- and post-treatment tests (65% vs. 70% correct, t(56)= 2.40, p=.01, one tailed). In contrast, for the Placebo Group recall of these words did not change across the two testing sessions (67% vs. 68% correct, t(58) = 0.47, n.s. A surprising feature of *Hopkins* task performance was that for both groups, recall of words from the first half of each list declined between pre-and post-treatment tests, F(1,114)= 8.79, p<.005. The extent of this decline did not vary between treatment groups, F<1. This finding was unexpected, and may reflect a shift in the strategy used by participants for recalling the word lists in the pre- and post-treatment testing sessions. That is, in the second testing session participants may have devoted fewer processing resources to recalling words from the first part of the list.

Separate multiple regression analyses were performed with respect to free recall of early and recent items from the *Hopkins* word lists (i.e. with respect to recall of words from the first and second halves of the lists). Initial screening of data prior to these analyses revealed an abnormally high centred leverage value for the observations from one participant. This was the result of an extreme outlier with respect to ferritin change. One individual in the Iron group underwent an increase in ferritin level that was dramatically larger than for her peers (nearly five standard deviations from the mean for this group). Following the recommendation of Tabachnick and Fidell (1996) the ferritin change score for this individual was reduced to a value one unit larger than the next most extreme score. This brought the leverage, Cook's D values, and studentised deleted residuals within acceptable boundaries, as described above. Since, undue influence from this outlier was also apparent in subsequent multiple regression analyses, this Winsorising procedure (affecting just one ferritin change score) was used in each of the multiple regressions reported below.

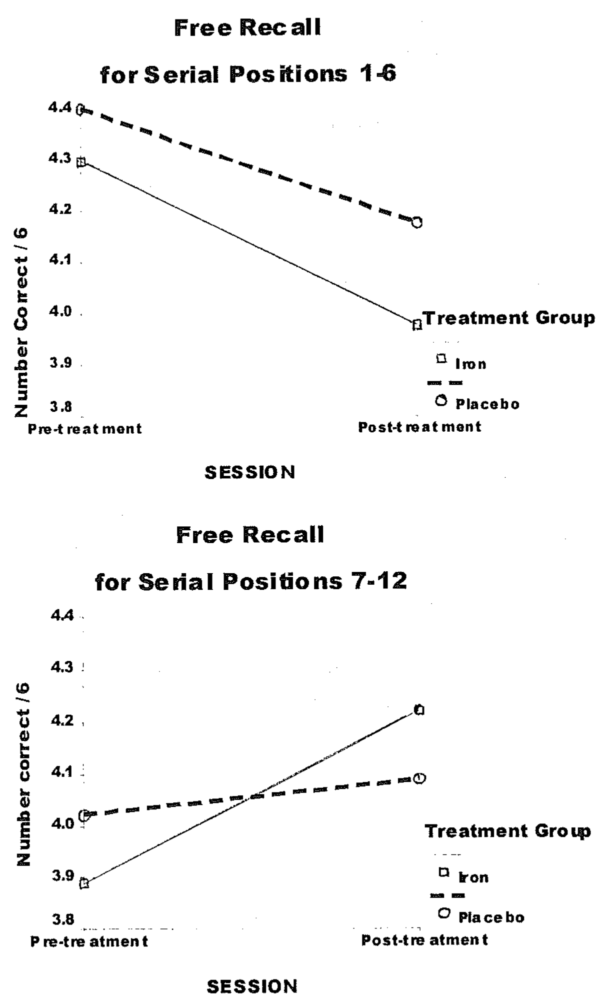
No significant relationships were observed between haematological measures of iron status and recall of words from the first half of the list.

The outcome of a multiple regression analysis examining recall of words from the second half of the list is summarised in Table 2. There was a significant (p<.004) relation between haemoglobin change and post-treatment performance, controlling for pre-treatment performance. That is, consistent with the working memory hypothesis there was an association between the amount by which haemoglobin levels increased and the relative improvement in free recall of recently heard words. At post-test the simple bivariate correlation between haemoglobin levels and recall of recently heard words was r =.142, p<.07 (one tailed). This might be interpreted as tentative evidence for a relationship between haemoglobin levels and working memory in the general population of female adolescents. However, if this interpretation is accepted, the value of the correlation obtained here (r =.142) is likely to underestimate the strength of the relationship between haemoglobin and task performance in the general female adolescent population. The reason for this is that the range of haemoglobin scores was restricted by definition within our iron-deficient participants. That is, compared to the general population of female adolescents there was a relative paucity of scores from the upper end of the distribution. Furthermore, after the two month treatment period haemoglobin levels did not increase significantly even among participants receiving iron treatment – presumably because, as explained earlier, this measure of iron status responds relatively slowly to treatment. The most likely consequence of this restriction in the range of scores in our sample is that measures of the correlation between haemoglobin and performance will underestimate the strength of relationships present in the population as a whole. An important issue for further research will be to examine haemoglobin – performance relationships in both iron-deficient and non-iron-deficient participants in order to investigate this possibility.

Table 2: Multiple regression of post-treatment immediate free recall (Hopkins Test: Serial Positions 7-12)

PREDICTOR VARIABLES	Standardised Coefficients (Beta)	t	p	Correlations: Bi-variate	Correlations: Semi-partial	R^2
Pre-treatment Performance	.32	3.63	<.001	.295	.317	.100
Age	.05	0.56	n.s.	.028	.049	-
Ferritin Change	-.03	0.30	n.s.	.063	-.026	-
Haemoglobin change	.25	2.75	<.004 (one tailed)	.225	.240	.058

Figure 1. Immediate free recall performance on the *Hopkins Verbal Learning Test*. The upper panel shows recall of early items in the list; the lower panel shows recall of recent items.



Delayed recall and recognition scores of the HVLT were also examined using split-plot ANOVA and multiple regression. As Bruner et al. (1996) also observed, there were no significant relationships between iron treatment and these measures of performance.

Reading Span Test: Mean reading span scores before and after treatment are shown in Figure 2. In a split-plot analysis of variance the only effect to approach significance was the main effect of Treatment Group, F(1,114)= 3.28, p<.08. It is noteworthy that the performance of the Iron Group was significantly better than the Placebo Group prior to the treatment phase, t(114)=2.16, p<.04, (two tailed). Since participants were assigned to the two groups randomly, we interpret this result as a Type 1 error. Unfortunately, this pre-treatment difference complicates interpretation of the data shown in Figure 2 in the following way.

Between the two testing sessions one would expect the data to be subject to two distinct and opposing influences. Firstly, since the pre-treatment difference almost certainly reflects Type 1 error, one would expect that the difference between the Iron and Placebo groups would regress towards zero – the expected value, assuming no treatment effect. As Figure 2 shows, the difference between the two groups does indeed reduce between pre-and post-treatment. However, since a treatment effect is hypothesised, we expect an additional influence, whereby the scores of the iron group improve relative to placebo. This would tend to increase the difference in scores between the two groups. According to this hypothesis, there should still be a difference between

Figure 2. Performance of Iron and Placebo groups on the *Reading Span Task.*

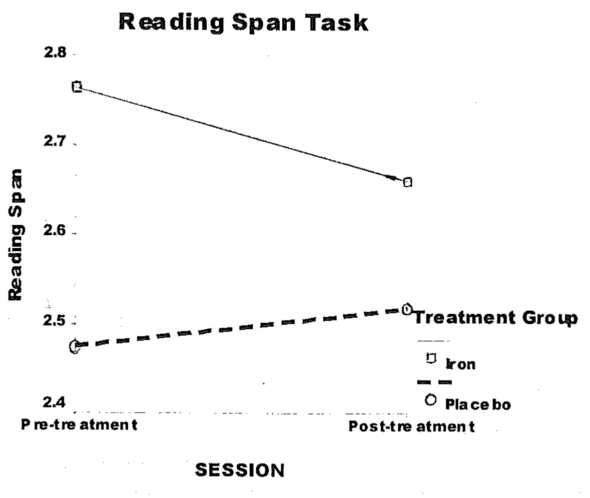


Table 3: Multiple regression analysis of post-treatment reading span scores

PREDICTOR VARIABLES	Standardised Coefficients (Beta)	t	p	Correlations: Bi-variate	Correlations: Semi-partial	R^2
Pre-treatment Performance	.40	4.64	<.001	.380	.394	.155
Age	-.13	1.51	n.s.	-.049	-.128	-
Ferritin Change	.22	2.46	<.01 (one tailed)	.184	.209	.044
Haemoglobin change	-.11	1.19	n.s.	-.014	-.101	-

the performance of the two groups at post-treatment, notwithstanding regression towards the mean. However, at post-treatment the difference in reading span between the Iron and Placebo Groups did not approach significance, t(114) =.94, n.s.

Reading span scores were also examined with multiple regression analysis, using the same format as described earlier. The outcome of this analysis is summarised in Table 3. As this table shows, consistent with the working memory hypothesis, there was a significant (p<.01) relationship between ferritin change scores and post-treatment reading span, while controlling for pre-treatment reading span. At post-test the simple bivariate correlation between serum ferritin and reading span was highly significant, r =.250, p<.004 (one tailed).

However, once again, this correlation may underestimate the strength of the relationship between ferritin levels and reading span performance in the female adolescent population generally. By definition, pre-treatment ferritin scores were all less than 12 mg/l. After the two month treatment period, the average ferritin levels of the Placebo Group remained in the iron-deficient range: 10.1 mg/l. Accordingly, the distribution of ferritin levels in this group was clearly abnormal, due to the relative paucity of scores from the upper end of the distribution. This was less true of post-treatment ferritin scores in the Iron Group. For the vast majority of these participants, ferritin levels did increase to a value within the normal range.

As Table 1 shows, the standard deviation and range of post-treatment ferritin levels and ferritin change scores were substantially larger for the Iron Group, in comparison with the Placebo Group. In view of this problem of restricted range in the sampling of post-treatment ferritin scores in the Placebo Group, a subsidiary analysis was performed, focusing on the relationship between ferritin and reading span within the Iron Group. This analysis employed the same predictor variables as shown in Table 3. The resulting semi-partial correlation between ferritin change and reading span was r = 0.325, yielding an R^2 value of 0.11. Within the Iron Group, the simple bivariate correlation between post-treatment ferritin levels and post-treatment reading span was r =.310, p<.01, one tailed.

Stroop Task: Naming times and number of naming errors for the Stroop Interference and Stroop Control conditions are shown in Table 4. Split-plot ANOVA and multiple regression analyses revealed no relationships of Stroop performance with treatment group or with haematological measures of iron status.

Table 4. *Stroop Task* Performance

TREATMENT GROUP	TESTING SESSION	*STROOP* NAMING TIME (SEC)	CONTROL NAMING TIME (SEC)	NAMING ERRORS
IRON GROUP	Pre-treatment	89.0	59.6	9.74
	Post-treatment	78.0	56.9	6.44
PLACEBO GROUP	Pre-treatment	89.3	61.4	9.61
	Post-treatment	76.9	59.7	7.39

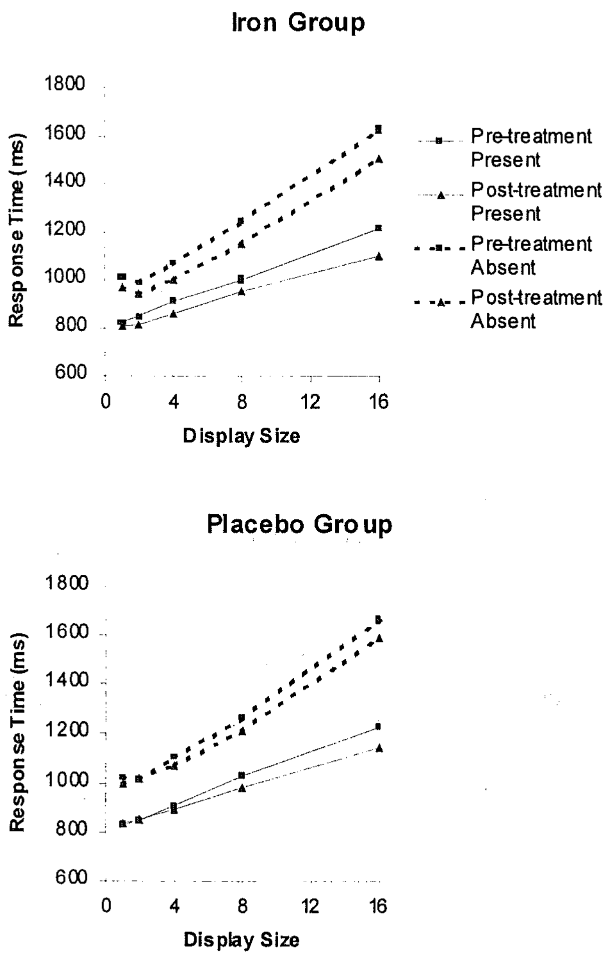
Visual Search task: *Visual Search* performance of the Iron and Placebo Groups is shown in Figure 3. These data were entered into a split-plot analysis of variance with the following factors: Testing Session Display Size (1, 2, 4, 8 or 16 items), Search Decision (Present vs. Absent) and Treatment Group. The predicted interaction between Treatment Group and Testing Session did not approach significance, F<1. The mean error rate across both testing sessions was 3.8%. In an analysis of variance of these data, there was no interaction between Treatment Group and Testing Session, F<1 (see Footnote 1).

Discussion

Results from this study present a somewhat mixed picture. Our primary analyses which tested for treatment effects (i.e. overall differences between the Iron and Placebo groups) provided only tentative evidence in support of the working memory hypothesis. On the one hand, some support for the hypothesis was gained from our first measure of working memory: Free recall of recently heard words improved between pre- and post-treatment testing sessions for participants in the Iron group, but not for those in the Placebo group (see lower panel of Figure 1). On the other hand, with respect to reading span, our second measure of working memory, the predicted interaction between treatment group and testing session was not observed (see Figure 2) However, as indicated earlier interpretation of these data is complicated by the presence of a spurious pre-treatment difference between the two groups.

However, this tentative support for the working memory hypothesis furnished by analyses of variance was bolstered by the outcome of MR and correlational analyses. These revealed: (1) An association between change in haemoglobin levels and the relative improvement in free recall of recently heard words; (2) An association between change in serum ferritin levels and relative improvement on the reading span task; (3) Significant correlation between post-treatment ferritin levels and post-treatment reading span performance.

Figure 3. Performance of the Iron and Placebo groups on the *Visual Search* task



As indicated earlier, the MR and correlational analyses can be seen as providing a more powerful test of the working memory hypothesis, since they make use of continuous information concerning precise changes in iron status, while the analyses of variance provides a dichotomous contrast between performance of the iron and placebo groups. Nevertheless, it should also be acknowledged that the picture provided by the current findings is less than straightforward. With respect to the clinically and practically relevant question of whether a group treated with iron supplements would outperform a placebo group, the data provided only tentative support for an affirmative answer.

Perhaps the best interpretation of our study is to propose that the data provide suggestive evidence in favour of two conclusions. The first is that non-anaemic iron deficiency may impair performance on everyday tasks that involve verbal working memory. That is, tasks involving the storage and processing of verbal information. This implies that iron-deficient individuals are likely to be disadvantaged relative to their peers, in terms of ability to perform everyday verbal tasks. The relatively high prevalence rates of iron deficiency both here in New Zealand (Schaaf et al., 2000) and world-wide (Scrimshaw, 1991) indicate that this is an important issue worthy of further investigation. In the context of New Zealand, where the study was performed, this conclusion also raises issues of ethnic equality, since, as mentioned earlier, iron deficiency appears to be especially prevalent among certain groups, especially those of Maori and Pacific Island ethnicity (Schaaf et al., 2000). Data from the New Zealand National Education Monitoring Project (Crooks & Caygill, 1999; see www.nemp.otago.ac.nz) suggest a picture in which students from these ethnic minorities under-perform across a variety of curriculum areas, relative to other students. This disparity in academic achievement probably arises from multiple causes, including socio-economic and cultural factors. However, the present results can be interpreted as suggestive evidence that nutritional factors may also contribute to low levels of academic achievement among female students of Maori and Pacific Island ethnicity.

The second conclusion, which may be seen as a corollary of the first, is that improved iron status may lead to to better performance on tasks that require storage and processing of verbal information. Clearly, improving iron levels could be achieved in at least two ways – by improvements in diet to prevent the onset of iron deficiency (primary prevention), or by screening for iron deficiency and giving iron supplements to those who are iron-deficient (secondary prevention). The cognitive benefits reported here suggest that secondary prevention strategies against iron deficiency would be successful in our study population, although detailed discussion of which prevention strategy is likely to be most effective as a public health measure is beyond the scope of this paper.

In view of the mixed nature of our results, further research is clearly needed to assess the soundness of these conclusions and to test for further relationships between iron deficiency and cognitive performance. Nevertheless, the practical implications of our findings, should they prove robust are considerable. Remember that all our adolescent participants were still attending school. Clearly, a wide variety of everyday school activities will rely heavily on the use of working memory. Indeed, it is difficult to conceive of any academically related tasks that would not involve working memory. The reading span measure of working memory was included in our test battery precisely because in previous research performance on this test has correlated strongly with measures of reading comprehension, including the verbal *Scholastic Aptitude Test* (Daneman and Carpenter, 1980, Masson and Miller, 1983).

Consistent with the results of Bruner et al. (1996) no relationships were found between iron treatment and delayed recall or recognition. However, this should not be taken as support for the assumption that iron status is unrelated to long term memory performance. The HVLT assesses memory with respect to storage of lists of unrelated words. In contrast, many everyday tasks involve memory for connected verbal information. In such tasks, impairment of working memory is likely to influence long term memory indirectly, through impairment of text comprehension, via processes of inference making and integration within working memory.

Neither of our measures of visual attention (*Stroop* or visual search) showed any relation with iron treatment. However, as noted earlier the notion of attention is multi-faceted, embracing concepts of selectivity, alertness and processing resources or effort. The latter aspect of attention is closely allied to the concept of working memory. Thus, while the results provide no support for the hypothesis that iron status is related to processes of selective attention, as we have seen they do provide support for the idea that iron levels are related to attention in the sense of processing resources within working memory.

In conclusion, results from the present study support the hypothesis that iron deficiency impairs performance of tasks involving verbal working memory, and improvements in iron status are associated with improved verbal working memory. However, it is acknowledged that our findings might be described as suggestive rather than conclusive. Accordingly, it is clear that further research into this important issue is required.

In future work it may be desirable to test for treatment effects over a longer period. In the current study it was clear that an eight week treatment period was insufficient to effect an improvement in haemoglobin levels. Similarly, it is conceivable that a longer period of iron treatment may be required to observe clear improvements in cognitive performance. In addition, as noted earlier, valuable information would be gained from further investigation of relations between iron status and working memory in both iron-deficient and non-iron-deficient individuals.

Another important issue concerns the relation between iron deficiency and other forms of working memory, for example involving the storage and manipulation of spatial or numerical information. The latter may be especially important, given that iron deficiency is particularly prevalent among females, and in view of the oft noted gender disparity in mathematical attainment at school.

References

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist, 5*, 125-142.

Bruner,A., Joffe, A., Duggan, A., Casella, J. & Brandt, J. (1996). Randomised study of cognitive effects of iron supplementation in non-anaemic iron-deficient adolescent girls. *The Lancet, 348*, 992-996.

Crooks, T. & Caygill, R. (1999). *New Zealand's National Education Monitoring Project: Maori Student Achievement, 1995-1998.* Paper presented at the Joint Conference of the Australian Association for Research in Education and the New Zealand Association for Research in Education (Melbourne, Australia, November 29-December 2, 1999). www.nemp.otago.ac.nz

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal learning and Verbal Behavior, 19*, 450-466.

Judd, C. M. & McClelland, G. H. (1989). *Data Analysis: A Model Comparison Approach.* Harcourt, Brace, Jovanovich: Orland, Florida.

Masson, M. & Miller, J. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology, 75*, 314-318.

Oski,F., Honig,A., Helu,B. & Howanitz,P. (1983). Effects of irontherapy on behavior performance in non-anemic, iron-deficient infants. *Pediatrics, 71*, 877-880.

Pollitt E. Soemantri AG. Yunis F. Scrimshaw NS. (1985). Cognitive effects of iron-deficiency anaemia [letter]. *Lancet. 1(8421)*:158,

Posner, M.I. (1994). *Images of Mind.* New York: W.H. Freeman (Scientific American Library).

Schaaf, D., Scragg, R., Metcalf, P. & Grant, C. (2000). Prevalence of iron deficiency in Auckland high school students. *New Zealand Medical Journal, 113*, 347-350.

Scrimshaw, N. (1991). Iron deficiency. *Scientific American, 265*, 46-52.

Tabachnick, B. & Fiddell, L. (1996). *Using Multivariate Statistics* (3rd Edition). Harper Collins: New York.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Lecture. *Quarterly Journal of Experimental Psychology, 40A*, 201-238.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

World Health Organisation (1972). *Nutritional Anaemias.* World Health Organisation technical Report Series, No. 503.

continues over page

Notes

1. In addition to these analyses, we also calculated the following measures of performance for each participant: (1) Rate of visual search function for present decisions. This was the slope of the best fitting linear regression between display size and mean response time for target present decisions. (2) Search intercept for present decisions – i.e. the intercept of the target present regression line. (3) Rate of visual search for absent decisions – i.e. the slope of the linear regression between display size and response time for target absent decisions. (4) Search intercept for absent present decisions – i.e. the intercept of the target absent regression line. Split-plot ANOVA and multiple regression analyses revealed no relationships between any of these performance measures and either treatment group or haematological measures of performance.

It was surprising that while on the one hand the interaction between testing session and treatment group did not approach significance, the pattern of means appeared consistent across experimental conditions. Careful inspection of Figure 3 reveals that for each of the five display sizes x two search decision conditions (Present vs. Absent), the mean reduction in response time across pre- and post-treatment testing sessions was greater for participants in the Iron Group than for those in the Placebo Group. For target present decisions, the mean reduction in response times between pre- and post-treatment for the Iron vs. Placebo groups, across the five display size conditions were: 14ms vs. –5ms, 40ms vs. –3ms, , 54ms vs. 17ms, 48ms vs. 45ms, and 112ms vs. 85ms, respectively. For target absent decisions, the corresponding values were: 48ms vs. 25ms, 46ms vs. –6ms, 68ms vs. 29ms, 90ms vs. 55ms, and 121ms vs. 73ms, respectively. That is, of ten possible comparisons between the two groups, the mean differences in all ten cases are consistent with the hypothesis of greater performance improvement in the Iron group relative to Placebo. However, the analysis of variance outcome indicates that this apparent consistency across conditions is completely swamped by inter-subject variability in the the amount of reduction in RT as a function of testing session. This, together with the absence of regression effects leads us to accept the null hypothesis in this case (i.e. that iron treatment has no effect on visual search performance).

Acknowledgments

This research was supported by the Health Research Council of New Zealand (Grant Number 99/047).

The authors would like to acknowledge the contribution of Abbott Laboratories (NZ), who provided the iron tablets (Ferrogradumet), Douglas Pharmaceuticals (Auckland), who manufactured the placebo tablets, and Diagnostic Medical Laboratory (New Zealand) who analysed the blood samples.

We would also like to acknowledge the contribution of Debbie Raroa and Lois Shaw who collected over 1,000 blood samples in the course of this study.

We would like to express our gratitude to the teaching and nursing staff at Auckland Girls Grammar for their cheerful co-operation and support for this work.

We thank Michael Hautus for valuable discussion on statistical matters.

Last, but certainly not least, we would like to thank the Auckland Girls Grammar School students and parents who agreed to take part in the research.

Address for correspondence:

Anthony Lambert
Department of Psychology
University of Auckland
Private Bag 92019
Auckland, NZ

Email: a.lambert@auckland.ac.nz