

Classwide Peer Tutoring with or without Reinforcement: effects on academic responding, content coverage, achievement, intrinsic interest and reported project experiences

CHEUNG CHUN CHUN & SAM WINTER, *Department of Education, University of Hong Kong, Hong Kong*

ABSTRACT *The present study examined the effects of Classwide Peer Tutoring (CWPT) under two conditions—with reinforcement (CWPT + R) and without reinforcement (CWPT – R)—on 77 students' spelling performance and intrinsic interest in lower secondary school Integrated Science. The students displayed educational attainments in the average range. Both CWPT approaches led to significant improvements in spelling test performance. Both groups also evaluated the programme positively. However, the CWPT + R group made significantly greater learning gains than did the CWPT – R students. These appeared to be related to higher levels of academic responding. Despite their impressive spelling gains, students in the CWPT + R group displayed significantly lower intrinsic interest in Integrated Science after involvement in CWPT. However, great caution should be exercised in interpreting this last result, as statistically significant effects were not found when group differences at post-test were examined by way of t-test and analysis of covariance methods. In summary, the data from this study suggest: (a) that the CWPT method of instruction is useful for helping lower secondary school students, (b) that reinforcers may increase learning gains by encouraging higher rates of responding, but (c) that there is an indication, subject to caution, that these reinforcers may possibly depress intrinsic interest in an academic activity (science in the case of this study).*

As its name suggests, Classwide Peer Tutoring (CWPT) is a peer tutoring system involving tutor–tutee pairs working together on a classwide basis. It typically involves careful selection of instructional content and materials, pairing of students for reciprocal tutoring, regular changes of partners, immediate error correction, points contingent upon performance, allocation of tutoring pairs into teams competing for highest point

totals, public posting of individual and team scores, and social rewards for winning teams. It is designed to accelerate student learning by increasing students' opportunities to respond and thereby increasing their levels of academic responding.

CWPT has been extensively researched. It has been found useful in producing gains in spelling performance among low-achieving students (Hall *et al.*, 1982; Maheady & Harper, 1987). Large-scale and long-term research has also been undertaken in this area (Greenwood *et al.*, 1987, 1989; Greenwood, 1991). Findings indicate that CWPT yields greater learning gains than traditional teaching methods (e.g. Greenwood *et al.*, 1989).

CWPT has been successfully extended to curriculum areas other than spelling. Areas covered include reading (Cook *et al.*, 1983; Otis-Wilborn, 1984; Greenwood *et al.*, 1989; Greenwood, 1991), assorted other language abilities (Greenwood, Dinwiddie *et al.*, 1984; Otis-Wilborn, 1984; Greenwood, 1991), and arithmetic and mathematics (Cook *et al.*, 1983; Greenwood, Dinwiddie *et al.*, 1984; Greenwood *et al.*, 1989; Greenwood, 1991).

CWPT learning gains have been explained in terms of heightened opportunity to respond and higher response rates (Delquadri & Greenwood, 1981; Hall *et al.*, 1982; Maheady & Harper, 1987). This point of view has been given support by research which demonstrates relationships between teaching arrangements (CWPT and others), student response levels and achievement outcomes (Greenwood *et al.*, 1984; Greenwood, Dinwiddie *et al.*, 1984).

While peer tutoring research in Hong Kong has generated a number of unpublished reports (Chan, 1987; Cheung, 1987; Chen, 1990; Cheung, 1993; Chan, 1995) and published studies (Winter, 1988, 1996; Pang, 1993), none involving CWPT has so far been reported. The present study afforded an opportunity to examine its applicability to an ethnic Chinese context.

CWPT lays great stress on a high level of positive reinforcement. In this respect, it reflects a similar tendency throughout the peer tutoring literature overall; see, for example, the early studies of Trovato and Bucher (1980) and Greer and Polirstok (1982) in reading and comprehension, and of Buckholdt and Wodarski (1978) in mathematics and reading. The high levels of immediate and positive reinforcement available through peer tutoring have often been seen as one of the features that distinguish it most clearly from ordinary classroom instruction (see e.g. a discussion by Goodlad & Hirst, 1989).

There is, of course, a vast body of literature confirming the powerful effects of reinforcement upon learning. Hattie (1992) found in a meta-analysis of 134 separate meta-analytic reviews (involving nearly 8000 studies and around 10 million students) that reinforcement generated an effect size of 1.13 upon academic achievement (i.e. a difference of 1.13 standard deviation units between experimental and control groups); the largest effect size for any factor examined, whether child-, teacher-, school- or instruction-centred. Furthermore, he reported a very low standard deviation (0.12) for this effect size, indicating that the effect of reinforcement upon learning is quite consistent across different studies.

It is perhaps because of this substantial effect of reinforcement that another more recent review, this time of interventions with exceptional children, found that behaviour modification programmes generated one of the largest effect sizes of all interventions examined (Forness *et al.*, 1997).

Notwithstanding this impressive learning effect, there remain questions about possible unwanted side-effects of reinforcement. Deci (1971) and Lepper *et al.* (1973) have

suggested that an 'overjustification effect' can operate, whereby a person's intrinsic interest in an activity may be undermined by inducing him or her to engage in that activity as an explicit means to some extrinsic reinforcer. Several studies appear to have confirmed the existence of such an effect (e.g. Deci, 1971; Lepper *et al.*, 1973; Greene *et al.*, 1976; Morgan, 1983; and Chung, 1995; the final study being with Hong Kong students). In each study, subjects who already were interested in an activity showed decreased interest after reinforcers were used and then withdrawn. In view of the heavy use of reinforcement in CWPT, and the possible damaging effects of reinforcement, it was decided that this study should collect pre- and post-test data on students' level of interest in the subject being tutored.

Several studies attest to the positive experiences reported by students involved in CWPT (Maheady & Harper, 1987; Greenwood *et al.*, 1987). While students' experiences have a validity in their own right, it is also useful to examine them in the light of students' initial expectations upon entering a project. Winter (1994) has demonstrated that students involved in peer tutoring sometimes come out of a project reporting experiences, which, though positive, are less positive than their initial expectations. He dubs this disparity a 'disillusionment effect', and argues that the extent of a student's disillusionment may influence his/her reactions to future involvement in a similar programme as much as, or more than, do his/her simple reported experiences. Accordingly, this study collected data on both expectations and experiences.

The study to be reported also provided a first opportunity to study the effects of CWPT when used with ethnic Chinese students. It also made possible an examination of the effects of high levels of reinforcement (as typically used in CWPT) upon learning. Further, the study allowed an examination of possible effects that reinforcement might have upon intrinsic interest and reported experiences (i.e. possible overjustification and disillusionment effects). Finally, the study incorporated collection of data on two process variables (academic responding and content coverage) as a way of identifying possible mechanisms underlying any learning, overjustification or disillusionment effects that might be observed.

Method

Subjects

The subjects in this study were ethnic Chinese students attending a Hong Kong secondary school. As a consequence of the mechanism by which school places are allocated to Hong Kong students, the school was catering to the needs of students in the 20th to 60th percentile range of ability. Teachers at the school delivered the curriculum in a mix of English, Cantonese (the local spoken dialect) and Chinese (the written form). The two Form 2 classes which constituted subjects in this study were among the lower achieving students within the school and so might effectively be considered to lie within the average range of ability. The two classes were of equivalent ability. Most, if not all, students in these classes were experiencing difficulty in spelling English words introduced in the Integrated Science curriculum. One class (20 females and 22 males) was assigned to be the CWPT + R (CWPT with reinforcement) group. These would participate in standard CWPT throughout the intervention. Another class (14 females and 21 males) was assigned to be the CWPT -- R (CWPT without reinforcement) group. The mean age of subjects was 13 years.

Design

The study incorporated both comparison group and multiple baseline (across stimuli) elements, insofar as there were two CWPT groups (one receiving high levels of reinforcement and the other receiving no programmed reinforcement at all), with each group undergoing seven successive phases of intervention, each involving three sessions and each involving a different set of words to be taught. The first and the second of the three sessions consisted of two 3-minute CWPT tutoring sessions. The third of the sessions involved a post-test spelling assessment for the words just taught, as well as a pre-test spelling assessment for the list of words to be taught in the next phase. The only exceptions to this rule were for Phase 1 (in which pre-test occurred at the start of the first session) and Phase 7 (the last phase, in which there was clearly no need for a further pre-test). Seven sets of pre-test and post-test spelling assessments were available for each of the two groups. Additional pre-test and post-test data, collected at the very start and end of the experiment, were collected on subjects' intrinsic interest in Integrated Science, and on their initial expectations towards, and final reported experiences of, the programme.

Procedure

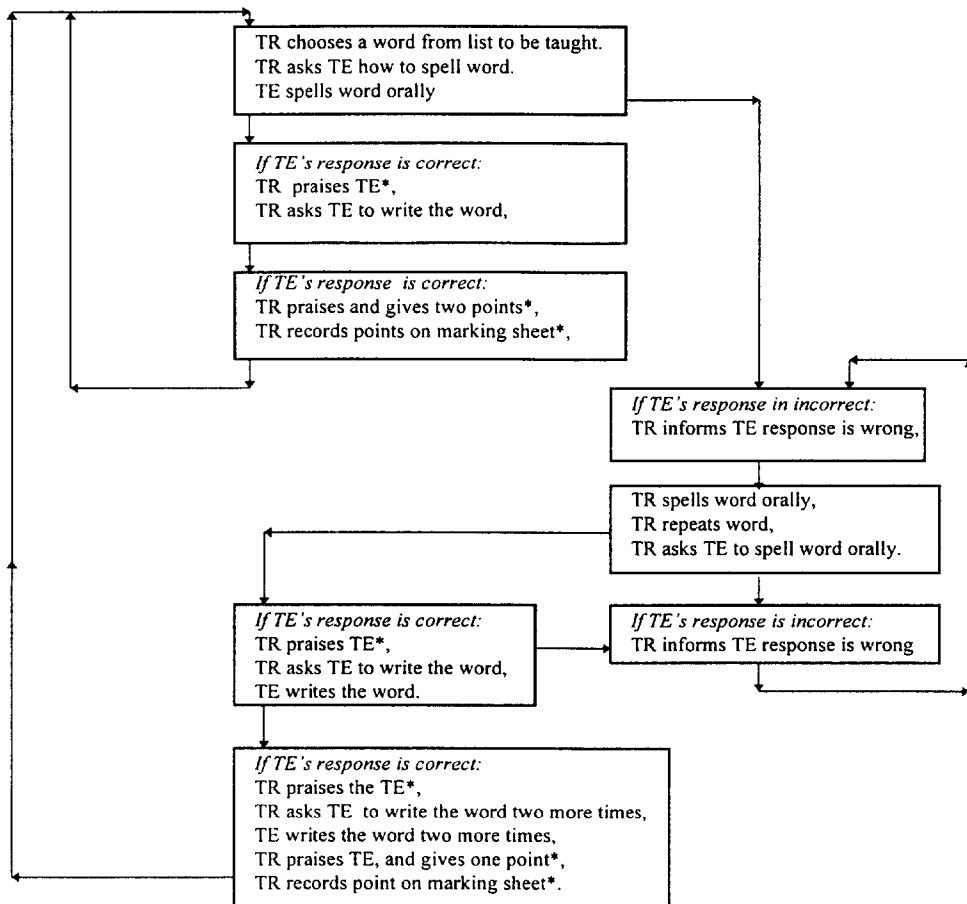
Before the start of the programme, students in both groups completed two questionnaires, delivered in the English language. The first questionnaire contained 10 items and measured students' intrinsic interest in the subject of Integrated Science. It was similar to one described by Linn and Gronlund (1995) and included items such as "I enjoy reading the textbook", "class activities are dull and boring", "I enjoy knowing more about science", and "the topics we study are unimportant", to which students were required to indicate agreement/disagreement on a five-point scale. The scale displayed a satisfactory level of internal consistency (alpha reliability coefficient of 0.82).

The second questionnaire contained five items and measured students' expectations regarding the programme in which they were about to take part. It was similar to one used by Winter (1994) and consisted of the following items: "I will be more confident in spelling", "I will spell more accurately", "I will spell faster", "Learning from a tutor will be more successful than learning from a teacher", and "Learning from a tutor will be more enjoyable than learning from a teacher." Students were required to indicate how likely they thought each of these things were, using a five-point scale running from 'very likely' to 'very unlikely'. The scale displayed a satisfactory alpha reliability coefficient of 0.77.

Subjects were trained in the use of CWPT procedures before the start of the experiment. Training proceeded by way of verbal instruction, practice through role-play and feedback. Subjects also received short booster training sessions after each of three short holidays that occurred during the study. English was used throughout the training sessions.

Students in the CWPT + R group were also placed in teams, who would compete for points during the project. Teams were changed regularly. CWPT - R students were not placed into any teams.

The tutoring procedures which CWPT + R students were trained to employ are shown in flowchart form in Fig. 1. CWPT - R teaching procedures were identical,



Note 1: TR = Tutor. TE = Tutee.

Note 2: * = Part of Teaching Format for CWPT+R group only.

Note 3: TR and TE exchange roles during each phase, with one session in each role.

FIG. 1. Teaching format for CWPT + R group.

except for the absence of tutors' reinforcement of tutee behaviour. Such reinforcement procedures are marked with an asterisk in Fig. 1.

Besides praise and points during each of the tutoring sessions, CWPT + R students enjoyed a wide range of other reinforcers, delivered at various points in the project. These were:

- (1) Bonus points reinforcement for subjects while in a tutor role: for example, for tutors who showed correct tutoring behaviours;
- (2) Public posting of points gained by individual students and teams;
- (3) Bonus points reinforcers contingent upon subjects' performance on post-test spelling assessments;

- (4) Certificate of accomplishments for a winning team, issued to each member of the team and publicly posted.

Two process variables were measured during each of the two tutoring sessions that took place in each phase of the study. Average values for these variables were calculated for each child in the two groups, as well as for each group. The two variables were:

- (1) The number of responses made by the subject in the tutee role (operationalised as the number of words written on a worksheet at the end of a session); and
- (2) The amount of content covered (operationalised as the number of words spelled correctly on the worksheet by the end of each session; with an upper limit of five—the maximum number of words taught—for each session).

In the absence of extensive checks on treatment integrity, these measures represented a check that students were actually engaging in spelling activities as prescribed by the teacher.

It was anticipated that there would be a close relationship, both metric and conceptual, between the two variables, with content coverage depending substantially on rate of academic responding, and with both acting as indirect measures of students' engagement in the programme.

For all seven word lists taught during the intervention (i.e. the seven phases of the study), data were collected regarding the number of words correctly spelled at pre-test and post-test. Average pre-test and post-test scores were calculated for each student, and for each of the two groups. These figures acted as outcome data for achievement.

Immediately after the seventh (last) phase of the study, students were required to complete two questionnaires again. One was the questionnaire on interest in Integrated Science, completed previously. The other was a questionnaire designed to assess subjects' subjective experiences of the programme. This was closely related to the expectations questionnaire that subjects had completed at the start of the study, except that the tenses were changed, and that students were now required to indicate the extent to which they agreed with the statements contained in the items (again on a five-point scale). Like the expectations questionnaire, it was based on one designed by Winter (1994).

Results

Spelling Test Performance

Pre-test spelling performance for the two groups was similar. The average pre-test score for the CWPT + R group was 5.02%, as against 3.03% for the CWPT - R group. A *t*-test confirmed that this difference was nonsignificant ($t = 1.73$, $df = 73$).

Both treatments appeared to lead to improved spelling performance. While the two groups combined were able to spell around 4% of words correct at pre-test, their performance at post-test rose to around 61% overall (see Fig. 2). The effect of CWPT seemed to be significant for each group.

A *t*-test (dependent samples, comparing pre-tests and post-tests) indicated that the CWPT + R group improved significantly ($t = -28.14$, $df = 39$, $p < 0.001$), with an average post-test score of 73.43% (a gain in score of 68.29 percentage points). A similar test revealed a significant gain for the CWPT - R group ($t = -14.98$, $df = 33$, $p < 0.001$), with an average post-test score of 48.40% (a gain in score of 45.38 percentage points of performance).

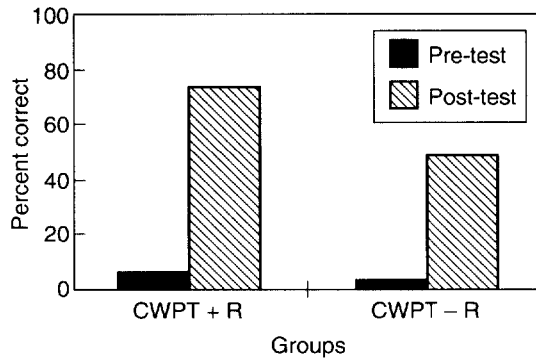


FIG. 2. Spelling performance.

The above figures suggest that, while there were significant spelling gains for both groups, those of the CWPT + R group seemed rather larger (by 22.91 percentage points of gain) than those of the CWPT - R group. A t -test (independent samples, examining the group difference in pre- to post-test gains) revealed that the group difference was significant ($t = 5.97$, $df = 72$, $p < 0.001$). Not surprisingly, the group difference in post-test scores was also statistically significant ($t = 5.97$, $df = 72$, $p < 0.001$).

In view of a sizable correlation between pre- and post-test performance (0.48, significant at $p < 0.001$), it was considered prudent to examine the post-test group difference in a way that took account of pre-test differences. An analysis of covariance, using post-test as the dependent variable and pre-test as the covariate, confirmed a significant effect for reinforcement in CWPT + R; the post-test difference was significant at $p < 0.001$ ($F = 31.06$). The analysis of covariance therefore confirmed that the incorporation of high levels of reinforcement into CWPT led to larger gains in spelling performance.

The impact of reinforcement upon CWPT achievement outcomes was significant in practical as well as statistical terms. The effect size for the group difference on post-test spelling scores was 1.34 (expressed in standard deviation units). A corresponding effect size of 1.30 was found when gain scores were used. These figures represent the practical effect of reinforcement in CWPT and compare favourably with the figure of 1.13 quoted by Hattie (1992) and discussed earlier in this paper.

Intrinsic Interest in Integrated Science

At pre-test, the mean interest scores for the CWPT + R and CWPT - R groups (3.63 and 3.44, respectively; with high scores indicating higher interest) were closely equivalent ($t = -1.56$, $df = 73$).

At first glance, CWPT appeared to have rather different effects on students' interest levels, depending on whether high levels of reinforcement were applied (see Fig. 3). The t -tests (dependent samples, to compare pre- to post-test changes for each group) suggested that the CWPT + R group displayed significantly reduced intrinsic interest in Integrated Science ($t = -2.38$, $df = 40$, $p < 0.05$). The drop was equivalent to 0.19 on a scale of 5. In contrast, the CWPT - R group displayed almost no change at all ($t = -0.21$, $df = 33$). It appears, therefore, that the use of reinforcement in CWPT had the effect of depressing intrinsic interest in the school subject involved. Notwithstand-

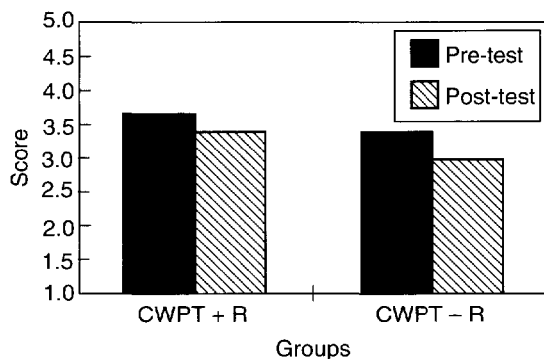


FIG. 3. Intrinsic interest.

ing the significant reduction in intrinsic interest for the CWPT + R students, the magnitude of the change (0.19 points) was really quite small (around 5% of the possible scoring range and only 0.36 pre-test standard deviations). Interestingly, this change was of the same magnitude as the initial pre-test group difference on this variable, which, owing to the nature of the distributions in question, was *not* significant! Accordingly, one needs to harbour caution in interpreting these findings.

Furthermore, the post-test difference in interest scores was very small indeed (0.02 on a five-point scale). Not surprisingly, then, a further *t*-test (independent samples, to examine the group difference in post-test scores) yielded a nonsignificant outcome ($t = -0.16$, $df = 73$; effect size 0.03). A similar analysis using change scores also generated a nonsignificant effect ($t = 1.35$, $df = 73$; effect size 0.29).

In view of the nature of this data (weighted scores from a 5-point scale), it was considered prudent to employ non-parametric tests as an additional way of analysing the data. Wilcoxon Matched-Pairs Signed Ranks tests were used for pre-post-test within-group comparisons, while Mann-Whitney U tests were used for intergroup comparisons. Results were consistent with parametric tests, indicating a significant pre-post-test change for the CWPT + R group, with all other comparisons being nonsignificant.

Finally, an analysis of covariance, undertaken to take account of the sizable correlation between pre-test and post-test interest (0.52, significant at $p < 0.001$) and the possibility that post-test differences were influenced by pre-test differences, confirmed the nonsignificant group effect ($F = 0.52$). In view of the fact that the analysis of covariance represents the most powerful of the tests used on this data, substantial doubt remains regarding the existence of an overjustification effect in this study.

Subjective Perceptions Regarding the Programme

Students in each group appeared to have quite positive initial expectations towards the programme (3.31 for the CWPT + R group and 3.42 for the CWPT - R group; with high scores indicating more positive expectations and experiences). This small difference was statistically nonsignificant ($t = 0.76$, $df = 73$). Scores for post-test experiences were 3.46 and 3.28, respectively.

A casual inspection of the data suggested that for each group there were disparities between expectations and experiences, with CWPT + R students reporting expecta-

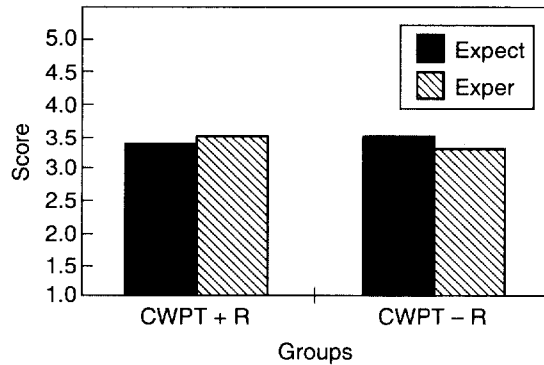


FIG. 4. Expectations and experiences.

tions that were less positive than their experiences and CWPT - R students reporting expectations that were *more* positive than their experiences (see Fig. 4).

The *t*-tests (dependent samples) were used to examine the significance of the disparities between expectations and experiences for each group. Results showed that they were nonsignificant for both the CWPT + R group ($t = 1.10$, $df = 40$) and the CWPT - R group ($t = -0.99$, $df = 33$).

As the observed disparities were in different directions, the disparity (i.e. change) scores themselves were examined for any significant difference. None was found ($t = -1.47$, $df = 73$). It was also considered useful to examine the significance of the apparent group difference in post-test experiences (2.54 for the CWPT + R group and 2.72 for the CWPT - R group). The difference was again nonsignificant ($t = -1.04$, $df = 73$).

Again, non-parametric tests were used as an additional way of analysing the data from these questionnaires. Wilcoxon Matched-Pairs Signed Ranks tests were used for pre-post-test within-group comparisons, while Mann-Whitney U tests were used for intergroup comparisons. Results from these analyses were consistent with those from parametric tests, with all comparisons being nonsignificant.

The correlation between experiences and expectations was quite small (0.23) and nonsignificant. Nevertheless, it was considered prudent to employ an analysis of covariance procedure using expectations as covariate and experiences as dependent variable. This procedure, like the others, produced a nonsignificant result ($F = 1.56$).

In summary, then, there was no reliable evidence of differences over testings for either group, or between groups at either testing. There was therefore no clear indication of a disillusionment effect for either group.

Process Variables

Both groups appeared to be highly engaged in the tutoring procedures (see Figs 5 and 6). Across the whole sample, students made an average 14.42 responses (as measured by the number of words written on a worksheet) over a 3-minute period; an average of nearly five responses a minute!

They also appeared to cover a large number of words (as measured by the number of words spelled correctly on the worksheet). Across the whole sample, they covered an average of 4.4 words per session (88% of the maximum possible; given that only five words were scheduled to be taught in each session).

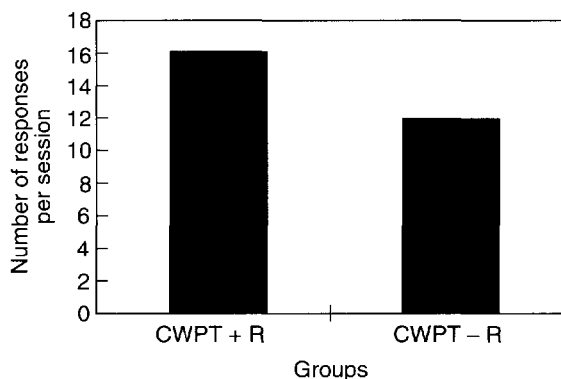


FIG. 5. Academic responding.

In the absence of more thorough checks on treatment integrity, these figures provide reassurance that the students were actually engaging in the spelling activities prescribed for the programme. Indeed, the high rate of responding seems, at face value, to indicate that the students were engaged in little else during the sessions.

Turning now to group differences, CWPT + R tutees appeared to make a somewhat higher number of responses than did those in the CWPT - R group (16.26 as against 12.59). They also appeared to cover more content (4.47 words against 4.35). Independent *t*-tests were used to examine the significance of these group differences in tutoring. The group difference for responses was significant ($t = 5.13$, $df = 72$, $p < 0.001$), with an average 3.66 more words written on CWPT + R worksheets than on the CWPT - R worksheets. However, the group difference for content coverage was nonsignificant ($t = 1.11$, $df = 72$). In summary, then, there was some evidence that students engaged in CWPT responded faster where high levels of reinforcers were being used, but not that they covered more content.

Correlation among Outcome Evaluation, Spelling Test Performance and Process Variables

For the overall sample and for each group separately, an examination was made of the degree of association between the process variables (number of responses and content coverage) and outcome (expressed as spelling gains, changes in interest, and disparities

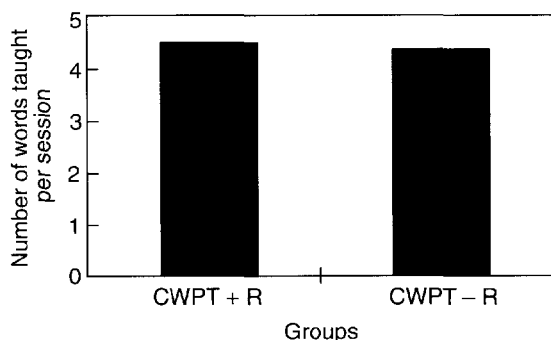


FIG. 6. Content coverage.

TABLE I. Correlations between process and outcome variables

Outcome variables	Process variables					
	Overall		CWPT + R group		CWPT - R group	
	Responses	Content coverage	Responses	Content coverage	Responses	Content coverage
Spelling gain	0.50**	0.38**	0.46*	0.33	0.03	0.40
Change in interest	0.11	-0.03	0.05	0.08	0.05	-0.18
Disparity between expectation and experiences	-0.01	0.09	0.13	0.23	0.03	-0.01

* $p < 0.01$; ** $p < 0.001$.

between expectations and experiences). The relevant Pearson correlation coefficients are displayed in Table I.

For the overall sample, and for the CWPT + R group, the correlations between responses and spelling gain were substantial in size, as well as statistically significant, indicating again that higher levels of responding were associated with larger spelling gains. For the overall sample, there was also a significant correlation between content coverage and spelling, indicating that it was the students covering more words who gained most from the programme. The corresponding correlations for each of the two groups fell just short of significance.

In summary, then, correlational analysis (using individual students as the units for analysis) provided additional evidence, over and above that at the group level, for the relationship between academic responding and achievement, and furthermore provided some evidence to suggest a link between content coverage and achievement.

There appeared to be no relationships at all between the process variables and either of the two other outcome variables (change in interest level and disparity between expectations and experiences).

Discussion

The outcome data from this study confirm the value of the CWPT instructional method with ethnic Chinese students. Significant pre-to post-test gains were made by each of the groups; findings which echo those of earlier studies into the use of this approach as a way of teaching spelling (Hall *et al.*, 1982; Greenwood, 1991).

Earlier researchers have noted that the gains under CWPT may be due to the favourable 'ecological' arrangements it provides, enabling students to engage in a greater amount of academic responding (e.g. Greenwood *et al.*, 1989; Greenwood, 1991). Informal observations during this study indicate that students in each group were indeed highly occupied during the tutoring sessions. The data on academic responding and content coverage, reported earlier, confirm this.

Group data from the study appear to provide supportive evidence for the role of *academic responding* in generating learning gains under CWPT. The CWPT + R group made an average 30% more responses per session than did the CWPT - R group; a difference that was statistically significant. They also learned nearly three-quarters of the words taught, against around half for the CWPT - R group (again, statistically significant). In the present writers' view, the most direct explanation for the group

differences in academic responding and learning gains is that *one group experienced much higher levels of reinforcement than did the other, that the reinforcement encouraged the higher rates of responding and that the high response rates produced the large learning gains*. Note in this connection the correlational data from this study, which appear to provide, at the level of the individual student, some support for the role of academic responding in generating learning outcomes.

By contrast, group differences in *content coverage* were small and failed to reach statistical significance. In retrospect, it seems likely that the failure to observe an effect might have been the result of a ceiling effect (average figures for both groups approached the maximum of five words possible in any one session). It should be noted that correlational analysis at the level of *individual subjects* suggested a link between content coverage and learning gains. A significant correlation was found between the two variables over the whole sample, while equivalent figures for the two groups viewed separately just failed to reach significance.

Notwithstanding the beneficial effect reinforcement appears to have upon response rate and learning, the study also provides an indication of possible detrimental effects upon students. Students in the CWPT + R group displayed significantly lowered interest in Integrated Science after involvement in the programme; an effect not found for the CWPT - R group. However, the magnitude of change was small, serving only to remove nonsignificant group differences at pre-test. Accordingly, extreme caution must be exercised in interpreting this finding as evidence for an overjustification effect; especially as an analysis of covariance test failed to reveal any difference between the two groups.

Even if one accepts that CWPT + R led to reduced interest in Integrated Science, there remains doubt regarding the underlying mechanism for such an effect. It may be that reinforcement leads directly to decreased interest (i.e. that it in some way debases a previously valued activity) as argued by Deci (1971) and others. Alternatively, it might have been the *increased academic responding* brought about by the reinforcement that led to the reduced interest. Put quite simply, the CWPT + R group may have ended the project less interested in Integrated Science because they had been working much harder on their spellings throughout the project and had consequently grown tired of the subject in a way that was not the case for the CWPT - R students. However, it should be noted that the low correlation between, on one hand, number of responses and, on the other hand, interest change scores (0.11 over the whole sample; see Table I) calls this account into question.

Turning now to students' evaluation of the programme, previous studies have indicated that students typically report positive experiences under CWPT (Greenwood *et al.*, 1987; Maheady & Harper, 1987). Students in both groups in the present study did indeed report positive experiences. Furthermore, data on disparities between expectations and experiences failed to support the existence of a disillusionment effect as reported by Winter (1994). While there was a worrying tendency for the CWPT - R group to report experiences less positive than their expectations, the effect was non-significant and was in any case outweighed by an opposite tendency (again, non-significant) in the CWPT + R group.

Many of the conditions listed by Winter (1994) as contributing to disillusionment (e.g. highly structured techniques, few opportunities for decision-making, unchanged tasks) were in place in this study, as they were in Winter's (1994) study. However, it may be that temporal aspects of the CWPT programme, shorter and less intensive than the one described by Winter (1994), generated fewer negative feelings than would have

been the case for a longer and more intensive one. Additionally, it may be that the constantly changing and difficult spelling lists provided sufficient novelty and challenge for students to ensure relatively positive feelings, even until the end of the study.

One of the interesting findings of this study was that there was no apparent link, either at the level of group data or at the level of individual subjects in each group, between the two process variables and subjects' reported experiences. It appears that these two aspects of tutoring had no effect whatsoever on students' perceptions of the programme. To the extent that there was no group effect on experiences, it appears that even the presence/absence of reinforcement has no effect on students' perceptions of the programme. As was the case for the Winter (1994) study, students' experiences of the project appeared to be influenced by their expectations and little else. Indeed, the correlation between the two was 0.46 over the entire sample (statistically significant at $p < 0.001$).

Summary

The findings of this study, the first involving CWPT with ethnic Chinese students, confirm the potential usefulness of a few minutes of daily CWPT, with or without reinforcement, as a vehicle for delivering instruction for moderate- and low-achieving lower secondary school students. Students make substantial learning gains and evaluate the approach positively. Higher levels of academic responding and larger learning gains are found where high levels of reinforcement are employed. The study also provides an indication that high levels of reinforcement during CWPT may *possibly* lead to depressed interest in the subject matter being covered.

However, the research reported in this paper (a) was short-term and small-scale, (b) involved a wide variety of reinforcers, each of which might have somewhat different effects upon learning and affect, and (c) focused only on a few outcomes (all of them very short-term) out of the many possible outcomes of CWPT. This study should therefore be regarded as preliminary only.

A further shortcoming of the study was that treatment integrity checks were not incorporated into the programme. While the process data show that the two groups differed in terms of the number of responses being made, data are noticeably missing with regard to whether tutors were using reinforcers as they were supposed to (awarding or withholding them, according to the group to which they belonged). A doubt therefore remains as to extent to which the two groups participated in divergent treatments.

Finally, the measure of intrinsic interest focused on students' interests in Integrated Science. Arguably, any overjustification effects for reinforcement may have been more pronounced on student interest in spelling or the English language (neither of which were actually measured in this study).

Further research needs to be conducted (particularly with Asian students, who have seldom acted as research subjects in either peer tutoring or overjustification research) to examine further:

- (1) The effects of reinforcement upon student attitudes, incorporating integrity checks to ensure participants are using reinforcers as directed, and with a particular emphasis upon the differential effects of alternative forms of reinforcement. In this regard, Chung (1995, working with Asian students) has argued that

extrinsic rewards are less damaging when they are used in a way that conveys explicit information about students' competence and self-determination.

- (2) The effects of longer and more intensive CWPT projects on learning gains and attitudinal outcomes, as well as the effects of various levels of practice and task difficulty within sessions.
- (3) The effects of CWPT on other areas of student and group functioning (such as attitudes towards education, self-esteem, peer relationships).
- (4) The long-term effects of CWPT on all of these outcomes (achievement-related, as well as affective).

Correspondence: Sam Winter, Associate Professor, Department of Education, University of Hong Kong, Pokfulam Road, Hong Kong. Fax: (852) 28579279. Email: sjwinter@hkusua.hku.hk

REFERENCES

- BUCKHOLDT, D.R. & WODARSKI, J.S. (1978) The effects of different reinforcement systems on cooperative behaviors exhibited by children in classroom contexts, *Journal of Research and Development in Education*, 12, pp. 50–68.
- CHAN, K.C.L.S. (1987) Paired Reading: a comparison of the effectiveness of student teachers and peers in the tutoring of poor Chinese readers in a primary school in Hong Kong, *Unpublished M.Ed. dissertation*, University of Hong Kong.
- CHAN, S.Y.S. (1995) An investigation of the effectiveness of cross-age peer tutoring on writing in a band 5 Anglo-Chinese school in Hong Kong, *Unpublished M.Ed. Dissertation*, University of Hong Kong.
- CHEN, E. (1990) Mildly mentally handicapped peers as tutors in a social skills programme for autistic children, *Unpublished M.Ed. dissertation*, University of Hong Kong.
- CHIUNG, C.Y.C. (1993) One to one cross-age peer tutoring and same-age peer tutoring in English dictation: a comparative study, *Unpublished M.Ed. dissertation* University of Hong Kong.
- CHIFUNG, M.K.T. (1987) Peer tutoring with hearing impaired special school pupils, *Unpublished M.Ed. dissertation*, University of Hong Kong.
- CHUNG, K.T. (1995) The effects of extrinsic reinforcement on intrinsic motivation amongst mildly mentally handicapped children, *Curriculum Forum*, 4, pp. 98–114.
- COOK, N.K., HERON, T.E. & HEWARD, W.L. (1983) *Peer Tutoring: implementing classwide programs in primary grades* (Columbus, Special Press).
- DECI, E. (1971) Effects of externally mediated rewards on intrinsic motivation, *Journal of Personality and Social Psychology*, 18, pp. 105–115.
- DELQUADRI, J. & GREENWOOD, C.R. (1981) *The Importance of 'Opportunity to Respond' in the Education of the Minority Student* (Lawrence, University of Kansas) (ERIC Document Reproduction Service No. EC 133–363).
- FORNESS, S., KAVALE, K., BLUM, I. & LLOYD, J. (1997) A mega-analysis of meta-analyses: what works in special education and related services, *Teaching Exceptional Children*, July/August, pp. 4–9.
- GOODLAD, S. & HIRST, B. (1989) *Peer Tutoring: a guide for learning by teaching* (London, Kogan Page).
- GREENE, D., STERNBERG, B. & LEPPER, M. (1976) Overjustification in a token economy, *Journal of Personality and Social Psychology*, 34, pp. 1219–1234.
- GREENWOOD, C.R., DELQUADRI, J. & HALL, R.V. (1984) Opportunity to respond and student academic performance, in: W. HEWARD, T., HERON, J. TRAP-PORTER & D. HILL (Eds) *Focus on Behavior Analysis in Education*, pp. 58–88 (Columbus, Charles Merrill).
- GREENWOOD, C.R., DINWIDDIE, G., BAILEY, V., CARTA, J.J., DORSEY, D., KOHLER, F.W., NELSON, C., ROTHOLZ, D. & SCHULTE, D. (1987) Field replication of classwide peer tutoring, *Journal of Applied Behavior Analysis*, 20, pp. 151–160.
- GREENWOOD, C.R., DINWIDDIE, G., TERRY, B., WADE, L., STANLEY, S., THIBADEAU, S. & DELQUADRI, J. (1984) Teacher- versus peer-mediated instruction: an ecobehavioral analysis of achievement outcomes, *Journal of Applied Behavior Analysis*, 17, pp. 521–538.

- GREENWOOD, C.R., DELQUADRI, J.C. & HALL, R.V. (1989) Longitudinal effects of classwide peer tutoring, *Journal of Educational Psychology*, 81, pp. 371–383.
- GREENWOOD, C.R. (1991) Classwide peer tutoring: longitudinal effects on the reading, language, and mathematics achievement of at-risk students, *Reading, Writing, and Learning Disabilities*, 7, pp. 105–123.
- GREER, R.D. & POLIRSTOK, S.R. (1982) Collateral gains and short-term maintenance in reading & on-task behavior by some inner-city adolescents as a function of their use of social reinforcement while tutoring, *Journal of Applied Behavior Analysis*, 5, pp. 123–139.
- HALL, R.V., DELQUADRI, J.C., GREENWOOD, C.R. & THURSTON, L. (1982) The importance of opportunity to respond to children's academic success, in: E. EDGAR, N. HARING, J. JENKINS & C. PIOUS (Eds) *Serving Young Handicapped Children: issues and research*, pp. 107–140 (Baltimore, University of Baltimore).
- HATTIE, J. (1992) Measuring the effects of schooling, *Australian Journal of Education*, 36, pp. 5–13.
- LEPPER, M.M., GREENE, D. & NISBETT, R.E. (1973) Undermining children's intrinsic interest with extrinsic reward: a test of the 'overjustification' hypothesis, *Journal of Personality and Social Psychology*, 28, pp. 129–137.
- LINN, R.L. & GRONLUND, N.E. (1995) *Measurement and Assessment in Teaching*, 7th Edn (Englewood Cliffs, NJ, Prentice-Hall).
- MAHEADY, L. & HARPER, G.F. (1987) A classwide peer tutoring program to improve the spelling test performance of low-income, third- and fourth-grade students, *Education and Treatment of Children*, 10, pp. 120–133.
- MORGAN, M. (1983) Decrements in intrinsic motivation among rewarded and observer subjects, *Child Development*, 54, pp. 636–644.
- OTIS-WILBORN, A.K. (1984) The evaluation of the effects of four reading instructional procedures on the achievement of hearing-impaired children, Lawrence, University of Kansas.
- PANG C.W.F. (1993) A paired reading project for low-achievers in Hong Kong, in: S. WINTER (Ed.) *Special Students in Mainstream Schools*, Education Papers No. 19 (Hong Kong, University of Hong Kong).
- TROVATO, J. & BUCHER, B. (1980) Peer tutoring with or without home-based reinforcement for reading remediation, *Journal of Applied Behavior Analysis*, 13, pp. 129–141.
- WINTER, S. (1988) Paired reading: a study of process and outcome, *Educational Psychology*, 5, pp. 135–151.
- WINTER, S. (1994) Tutor supervision, tutoring behaviors and outcome in peer-tutored paired reading, *Unpublished doctoral dissertation*, University of Hong Kong.