# The Efficacy of Written Teacher Advice (Tip Sheets) for Managing Classroom Behaviour Problems

Emma Little , Alan Hudson & Ray Wilks

# The Efficacy of Written Teacher Advice (Tip Sheets) for Managing Classroom Behaviour Problems

**EMMA LITTLE, ALAN HUDSON & RAY WILKS,** *RMIT University, Bundoora, Australia*

ABSTRACT     *This study evaluated the efficacy of a set of tip sheets developed to provide written advice to teachers about the management of common classroom behaviour problems. Twenty teachers were involved in the trial of the tip sheets. Teachers selected a child in their grade who demonstrated one of the behaviour problems, and after a baseline period, implemented the written advice for a period of 4 weeks. Outcome measures included teacher collected data, Goal Achievement Scale scores, Sutter–Eyberg Student Behavior Inventory, Teacher Report Form and consumer satisfaction. Measures of treatment integrity were also collected. Results indicated that 13 of the children showed substantial or moderate improvement. Of the seven who did not improve, four had high levels of general behaviour problems prior to intervention, and three had teachers who were not following the written advice completely. However, high levels of general behaviour problems did not prevent some other children still demonstrating improvements in behaviour (n = 5). The written advice in the tip sheets was found to be highly acceptable to the teachers.*

Conduct problems in the form of Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) constitute the most common reason for referral of children to inpatient and outpatient clinical facilities (Robins, 1991; Eme & Kavanaugh, 1995) with research suggesting they constitute more than 50% of referrals to clinical settings (Kazdin, 1989). Furthermore, conduct problems have wide ranging effects on family functioning, academic behaviour, adolescent behaviour and even adult psychopathology. In addition, when conduct problems occur in multiple settings, the effects are even more damaging in terms of later delinquency and substance abuse problems (Loeber *et al.*, 1991). The term conduct problems is used to describe externalising behaviours that are serious enough to interfere with a child's day to day functioning, with ODD and CD being the most severe forms of conduct problems.

Treatment of conduct problems in the home setting has received much attention in the literature, and there are well established, empirically derived treatment programmes that have been demonstrated to be effective. The hierarchical model of Behavioural Family Intervention developed by Sanders (1999) is empirically supported and is based on a minimal intervention/prevention model, which allows the clinician to select the appropriate level of intervention for individual families to promote change.

However, treatment for conduct problems in the classroom has not received a comparable amount of attention, and the intervention programmes are diverse, occasionally lacking empirical support and often not consistent with strategies used in the home setting. As past research has demonstrated that conduct problems in multiple settings are related to poorer prognosis, it is logical to suggest that interventions should focus on as many of the settings as possible in which a child displays the behaviour problems. As children spend most of their time outside of the home at school, this is one of the most important intervention settings for behaviour problems.

A recent Australian study indicates that 5.6% of children do demonstrate behaviour problems across settings (Little *et al.*, 2000). Therefore, for these children a consistent model of intervention across home and school is needed. Having hierarchical models of intervention for both settings that are based on the same theory and consist of similar strategies may increase the likelihood of generalisation across settings.

Researchers have also reported that one strategy for addressing behaviour problems will not 'fit' all young people (Keller & Tapasak, 1997). Therefore, a flexible approach to intervention is needed that allows strategies to be matched with the needs of the child. Not all children are the same in the extent to which they demonstrate behaviour problems across settings and as a result interventions need to be flexible across settings.

As already mentioned, a hierarchical intervention model for behavioural problems has been developed for the home setting (Sanders & Dadds, 1993). This model provides levels of intervention that increase in intensiveness. The lowest level (level 1) intervention for parents includes the use of Parent Tip Sheets (Department of Human Services, 1996). An evaluation of these tip sheets has been undertaken with results indicating that they are effective (Sultana *et al.*, 2000).

Given that written advice for parents on managing children's behaviour problems has demonstrated efficacy, the use of a similar approach with teachers would seem well worth investigating.

## Conduct Problems in the Classroom

Past research has found that the most common behavioural problems reported by primary school teachers are 'talking out of turn' and 'hindering other children' (Wheldall & Merrett, 1988). The other behaviours that were also identified as being troublesome included 'disobedience', 'idleness/slowness', 'making unnecessary noise' and 'aggression' (Wheldall & Merrett, 1988).

Similarly, research has found that children with conduct problems demonstrate difficulties in the classroom with listening, remaining on-task, attending, remaining in their seat, and answering questions (Hops & Cobbs, 1974; Shinn *et al.*, 1987). These behaviour patterns are ones that may occur quite frequently during class time and may result in the teacher spending a significant amount of time managing them. Given that these are associated with children demonstrating conduct problems, then it is reasonable to suggest that remediating these disruptive behaviour patterns may be an initial step in preventing the child from continuing down the conduct problem pathway

leading to fully diagnosable disorder (Conduct Disorder or Oppositional Defiant Disorder). Giving teachers step by step instructions on how to manage these problem behaviours may allow them to reduce the amount of time spent in managing behaviour problems, and increase the amount of time spent in instructional activities and in positive teacher-student interactions.

Minor conduct problems of children, such as those mentioned by past researchers (Hops & Cobbs, 1974; Wheldall & Merrett, 1988) could be managed by the routine and systematic use of reactive strategies by the teacher. Written materials similar to the 'tip sheets' in Behavioural Family Intervention could be used to advise teachers about how to respond systematically and effectively to commonly occurring classroom problems.

While gathering efficacy data is important in determining how successful an intervention is, it is as important to collect information about how well the intervention was implemented. Treatment integrity, sometimes called treatment adherence, refers to the extent to which an intervention has been carried out correctly and in accordance with the intention of the treatment designer (Armstrong *et al.*, 1997). In measuring treatment integrity, the consultant is actually observing how closely the consultee adheres to the set procedures and depending on the nature of the intervention, the consultant may then provide feedback to improve adherence.

Treatment Integrity is important in the examination of the efficacy of interventions as it allows the consultant to make assumptions about the extent to which behaviour change is actually due to the treatment protocol. It is of prime importance that the consultant is confident the intervention is carried out in the manner in which it was designed, particularly when another person has been trained to implement the procedures (Keller & Tapasak, 1997). If it cannot be demonstrated that the treatment was implemented correctly, then it is entirely possible that any change in the child's behaviour could be due to factors extraneous to the treatment programme. While researchers have repeatedly documented the need for treatment integrity measures, it is surprising to find that in intervention studies treatment integrity data are rarely collected. In a review of 233 studies on child and adolescent psychotherapy it was found that only 19% included information about treatment integrity (Kazdin *et al.*, 1990).

The efficacy of teacher consultation and intervention should theoretically be a function of how accurately and consistently the strategies are implemented by the teacher. It is then important that treatment integrity be measured in order to relate this to treatment efficacy. There are very few studies in the literature that address the effect of treatment integrity in psychological interventions on treatment outcome (Hogue *et al.*, 1996).

The aim of this study was to measure the efficacy of the teacher tip sheets for managing children's behaviour problems in the classroom setting. This aim covers two facets of efficacy. First, does the child's behaviour improve following the implementation of the strategies in the teacher tip sheet? Secondly, are the changes (or lack of) related to the degree that teachers adhere to the treatment protocol? It was hypothesised that:

(1) the use of teacher tip sheets would result in an improvement in the children's target behaviour;
(2) improvements would be related to treatment integrity;
(3) the tip sheets would be highly acceptable to teachers.

## Methods

*Participants*

Principals at Melbourne metropolitan schools ($n = 30$) were approached to have their teachers participate. The schools were selected on the basis of being within a 15 km radius of the university. If the principal agreed to his/her school being involved ($n = 5$), teachers who had a child in their class with one of the seven behaviour problems were invited to participate. A total of 21 Victorian primary school teachers agree to participate, representing four schools. There were 19 female teachers and two male teachers, which reflects the ratio of males and female teachers at the primary school level. The teachers ranged in age from 25 to 50 years. All primary school grade levels were represented. The schools represented low to high SES levels according to the Socio Economic Index for Areas programme (Australian Bureau of Statistics, 1992). One school was rated as low SES, two schools as middle SES and one school as high SES. The teachers had at least one child in their class with one out of the seven behaviour problems addressed by the teacher tip sheets. None of the children identified by the teachers who agreed to participate in the study were receiving assistance from another agency. One teacher dropped out of the study after baseline data collection.

*Teacher Tip Sheets*

The teacher tip sheets covered seven commonly occurring classroom behaviour problems:

- talking out of turn;
- hindering other children;
- having poor social skills;
- aggression;
- withdrawn behaviour;
- problems staying 'on-task';
- being disorganised.

Each tip sheet had a similar format that consisted of an introductory paragraph describing the problem behaviour, and the impact it has on the child and others. A brief paragraph then followed discussing possible reasons for a child developing such a problem behaviour. Strategies for prevention were then discussed, followed by strategies for when the problem behaviour occurs in the classroom. The suggested strategies were consistent with an applied behaviour analysis conceptualisation of classroom behaviours. Finally, a list of key steps summarising the information was provided for easy reference. Each tip sheet was two sides of an A4 sheet.

The acceptability of these tip sheets had already been established with a group of 60 teachers and 36 student teachers prior to the commencement of the efficacy study. Acceptability was measured using the Abbreviated Acceptability Rating Profile (Tarnowski & Simonian, 1992), which has eight Likert type items, each with a range of 1 to 6. Hence, the total acceptability score ranged from 8 to 48. A score of 32 or higher indicated that the mean rating by the teacher was at the acceptable end of the scale. Both teachers and student teachers rated the tip sheets as being acceptable with a mean acceptability score of 41.3 for teachers and 39.6 for student teachers.

*Measures*

*Treatment integrity measure.* Treatment integrity was measured through the use of classroom observations during both baseline and intervention phases. Each tip sheet had a corresponding integrity rating form that listed the teacher behaviours recommended in the tip sheet. The behaviours were divided into two categories: preventative strategies and reactive strategies. Preventative strategies are those behaviours that the teacher can use in order to lessen the likelihood of the child demonstrating the inappropriate behaviour (e.g. establishing rules, gaining eye contact when giving instructions, and praising appropriate behaviour). Reactive strategies are teacher behaviours that occur following the child's inappropriate behaviour (e.g. providing an appropriate consequence).

*Teacher Collected Data.* The teacher collected data form was designed by the researchers for the classroom teacher to record the behaviour problem as it occurred. The form divided the school day into four time periods (early morning, mid-morning, early afternoon and late afternoon). The teacher recorded each incident of the behaviour with a mark in the appropriate time period. At the end of that time period (e.g. at the start of recess), the teacher tallied the number of marks in that time period. At the end of the day a total frequency count was calculated by adding the scores for each of the time periods.

*Goal Achievement Scale.* The Goal Achievement Scale (Hudson *et al.*, 1995) was used to determine the clinical significance of the change following a behavioural intervention. This involved determining a baseline level of the behaviour, which was allocated the 0% success mark. Then, a decision was made as to what constituted 100% success. This decision was made in consultation with the teacher and was made following the baseline period. Given that the teacher was involved in the decision-making process as to the level of behaviour change required for 100% success, this potentially increased the likelihood of social validity of the intervention (Hudson, 1998).

The baseline level was calculated by averaging the number of times the child exhibited the inappropriate behaviour across all baseline days. The final level of behaviour achieved was calculated by averaging the last five data points from the last 5 days of intervention.

*Sutter–Eyberg Student Behavior Inventory.* The Sutter–Eyberg Student Behavior Inventory (SESBI, Sutter & Eyberg, 1984) is a 36-item inventory of conduct problem behaviours that are observable by teachers in the classroom situation. The aim of SESBI is to discriminate between 'normal' children and children with conduct disordered behaviours (Funderbunk & Eyberg, 1989). The SESBI is a narrow band measure of children's behaviour in the classroom. The SESBI rates each of the items of children's behaviour on two scales, intensity and its status as a problem. The Intensity Scale indicates how often the behaviours occur currently, and the Problem Scale indicates the specific behaviours that are problems currently for the teacher (Eyberg, 1992). The intensity Scale is rated using a frequency-of-occurrence rating ranging from 'Never' (1) to 'Always' (7), with a Total Intensity Score ranging from 36 to 252. Each item on the Problem Scale is rated either 'Yes' or 'No' depending on whether the teacher identifies that behaviour as currently being a problem for the target child. The

Total Problem Score ranges from 0 to 36. This measure has adequate levels of reliability and validity (Funderbunk & Eyberg, 1989; Schaughnency *et al.*, 1989).

Eyberg (1992) supports the use of the SESBI in measuring treatment gains. Eyberg reports that no evidence of regression or developmental effects has been found with the SESBI.

*Teacher Report Form.* The Teacher Report Form (Achenbach & Edelbrock, 1986) consists of 118 items that cover problem behaviours, of which 93 are comparable to items on the Child Behavior Checklist (Achenbach & Edelbrock, 1983). The other 25 items cover school related behaviours. Normalised T-scores are provided that compare the child to a random sample of same age and same sex peers (McConaughy *et al.*, 1988). Only scores on the Externalising and Total Problem scales were used in this study.

*Consumer Satisfaction Questionnaire.* The measure of consumer satisfaction was developed by the researchers and included seven questions, with some items based on the consumer satisfaction questionnaires developed by Eyberg (1974). The teacher completed this measure on termination of treatment. Each question was answered on a five-point (or six-point for two items) Likert scale (with a high rating being associated with a more positive response in all except one question). The teacher was also given an opportunity to make qualitative comments about the treatment programme.

### Procedure

*Teacher Selection of Target Behaviour.* Once a teacher agreed to participate in this study, she/he was given a list of the seven behaviours that were addressed by the tip sheets. The teacher then selected one behaviour that was problematic in a child in her/his classroom. The distribution of selected behaviours was as follows:

- eleven teachers selected 'talking out of turn';
- one teacher selected 'hindering other children';
- two teacher selected 'withdrawn' behaviour;
- one teacher selected 'aggression';
- six teachers selected 'off task' behaviour.

'Poor social skills' and 'disorganised behaviour' were not identified by any teacher as being of primary importance in intervention. Therefore, no data are available on the efficacy of those two teacher tip sheets.

*Collection of Baseline Data.* Ten teachers had a one week baseline (due to time constraints in the academic year) and 10 had a 2-week baseline. The first author (referred to as 'consultant' for the remainder of this paper) met with each teacher prior to commencement of baseline and at that time trained the teacher to complete the Teacher Daily Record Form. The consultant and the teacher decided on the observable behaviour/s that were to be measured and provided an operational definition of the target behaviour. The consultant also gave the teacher the SESBI and the Teacher Report Form to complete during the baseline period. The teacher then commenced 1 or 2 weeks of baseline data collection. The teacher then recorded every instance of the inappropriate behaviour for each school day on the Teacher Daily Record Form.

*Implementation of the Tip Sheet Advice.* Following baseline data collection, the Goal Achievement Scale was developed and the consultant gave the teacher a copy of the specific teacher tip sheet, which they read through together. The consultant clarified any areas the teacher was unclear on and gave the teacher examples of the strategies to be used following the target behaviour. The consultant then provided examples of the inappropriate behaviour and the teacher was required to demonstrate the appropriate response. This initial training session typically took 40–60 minutes. The consultant instructed the teacher to refer to the tip sheet on a regular basis (at least once per week). The teachers were instructed to keep the tip sheet on or near their desk (out of the view of students). The teacher then began using the intervention strategies with the child in his/her class.

Phone contact was made after the first week of treatment to monitor data collection and to answer any questions from the teacher. A second meeting between the consultant and the teacher was arranged at the end of the second week of intervention. In this second meeting the consultant collected the Teacher Daily Record Forms for the previous 2 weeks and briefly went through the strategies on the tip sheet with the teacher to ensure the teacher understood the procedures. This second meeting typically took 20–30 minutes.

The teacher continued intervention and data collection for a further 2 weeks. At the end of this 2-week period, the teacher and consultant met again. The consultant collected the Teacher Daily Record Forms and provided the teacher with the post-treatment measures (SESBI, TRF and Consumer Satisfaction Questionnaire).

For 10 teachers, a 6-week follow-up was conducted where they completed the Teacher Daily Record Form for 1 week. The other 10 teachers were unable to complete follow-up because of the end of the academic year.

*Assessment of Treatment Integrity.* Treatment integrity was measured through classroom observations conducted on three separate occasions across baseline and intervention (1 hour per observation). Treatment integrity was measured in two parts:

- preventative strategies;
- reactive strategies.

The integrity of the preventative strategies was calculated by recording the percentage of intervals during the observation period in which the teacher demonstrated those behaviours. As some of the strategies were specific to certain aspects of teaching (e.g. gaining eye contact when instructing the children), these could be scored as 'Not Appropriate' if not required (e.g. during individual seat work). The treatment integrity for the reactive strategies was measured by recording the teacher behaviours following the inappropriate behaviour occurring.

In computing a treatment integrity score, the mean percentage of correct behaviours for both preventative and reactive strategies was calculated. The mean of these two figures was then determined. If the integrity measure fell within the 0–50% range this was scored as low integrity. Scores from 51–75% were in the medium integrity range, and 76–100% were scored as high integrity.

*Data Analysis*

*Reliability of Teacher Collected Data.* In order to determine the reliability of the teacher's report of child behaviour, an independent observer conducted reliability checks during

the aforementioned treatment integrity checks. Two independent observers were used for the treatment integrity/interater reliability observations. The observers were trained in the use of both the treatment integrity measure and interater reliability measure.

Prior to the observation, the observer contacted each of the teachers and clarified the behaviour that was to be observed. The observer asked the teacher the specific behaviour/s that constituted the target behaviour and had the teacher provide examples. This ensured that the observer and the teacher were actually recording the same behaviours and reduced the subjectivity of the observations.

The observer then conducted the observation during part of the regular teaching day. The observer was seated in an unobtrusive position in the classroom where she could easily observe the target child and teacher. The teacher was asked to record the number of times the child displayed the inappropriate behaviour during the lesson. The observer also recorded each incident of the inappropriate behaviour that she witnessed. At the end of intervention, three sets of reliability data were available for each teacher.

A reliability coefficient was used in this study to determine inter rater agreement. A reliability coefficient was calculated for each set of data for the teachers in this study. Minimum, maximum and mean values were then calculated for the group.

*Graphing of Teacher Collected Data.* The average number of times the child demonstrated the inappropriate behaviour per session was graphed for each day of the study by the consultant. The average for each session was calculated because not all teachers had the same number of teaching sessions in a day and many teachers did not see their class for the same number of sessions each day of the week. It was decided that an average rate would be used to allow comparisons across teachers and across days.

*Rating of Graphs of Teacher Collected Data.* Three independent raters then visually inspected the graphs and rated the level of success. Each graph was rated by the independent raters using a 4-point rating scale: *substantial improvement*, *moderate improvement*, no change or deterioration. Substantial improvement was defined as 'data showing that following intervention there was an elimination of the inappropriate behaviour or a reduction to a very low level of occurrence' (given a rating of 3). Moderate improvement was defined as 'data showing that following intervention there was a clear reduction in the inappropriate behaviour but not sufficient to be considered substantial' (given a rating of 2). No change was defined as 'data showing that following intervention there was no change in the inappropriate behaviour' (given a rating of 1). Finally, deterioration was defined as 'data showing that following intervention the inappropriate behaviour was worse than at baseline' (given a rating of 0). When there was a discrepancy in ratings, the rating that two raters agreed upon was allocated to the graph.

*Global Ratings of Behaviour.* Pre and post intervention scores were collected for the SESBI Intensity and Problem Scales, and for the TRF Total and Externalising Scales. A repeated measures multivariate analysis of variance (MANOVA) was conducted on these variables to determine the significance (if any) of the behaviour change.

*Consumer Satisfaction Scale.* As previously described, the consumer satisfaction scale had four items with Likert scales (with a range of 1–5), and two items with Likert scales with six points (1–6). The sum of the six items was calculated for each teacher

providing a consumer satisfaction score with a maximum score of 32. Descriptive statistics were produced for the consumer satisfaction results.

*Relationship Between Measures.* In order to examine the relationship between treatment integrity, pretreatment global ratings of child behaviour and outcome (GAS or ratings of graphs of teacher collected data) multiple regression analyses were conducted. Independent regression analyses were conducted for the two outcome measures. The dependent variables were the GAS scores (0–100) and ratings of graphs of teacher collected data (0–3). The independent variables were pretest scores on the SESBI and TRF, and Treatment Integrity measures.

## Results

Of the 21 teachers who were originally involved in this study, one teacher commenced baseline and then decided that she could not continue with the intervention due to time constraints. Therefore, data are only available from 20 of the teachers.

First, the data on treatment integrity are presented. Secondly, an analysis of the teacher collected data will be presented. Thirdly, the Goal Achievement Scale scores will be examined. Fourthly, the data from the global ratings of child behaviour will be presented. Fifthly, data of consumer satisfaction will be presented. Finally, the relationship between measures will be examined.

### Treatment Integrity

At baseline, 18 of the 20 teachers had integrity scores in the low range (0–50%), two in the medium range (51–75%), and none in the high range (76–100%). During intervention only four of the 20 teachers had treatment integrity scores in the 'low' range. Of the remaining 16 teachers, seven had treatment integrity scores in the 'medium' range and nine in the 'high' range.

The relationship between the treatment integrity measures before and during intervention was examined using a repeated-measures $t$-test. A significant difference was found between the preventative strategies integrity score (baseline mean $= 36.1$, intervention mean $= 73.9$), $t(19) = 5.329$, $P < 0.001$. A significant difference was also found between the reactive strategies integrity score (baseline mean $= 11.5$, intervention mean $= 66.7$), $t(19) = 6.842$, $P < 0.001$. This indicates that the teachers as a whole demonstrated a significant improvement in the appropriate use of the treatment strategies from baseline to intervention.

### Teacher Collected Data

*Reliability of Teacher Collected Data.* The Pearson product-moment correlations between teacher/observer ranged from $r = 0.5$ to $r = 1.0$ (mean $= 0.96$). Only three of the 20 pairs had correlation coefficients less than 0.9. This indicates that inter rater reliability of the Teacher Daily Record Form was high.

*Reliability of Ratings of Graphs of Teacher Collected Data.* The three independent raters demonstrated a high level of agreement when determining the level of change demonstrated in the graphs of child behaviour (deterioration $= 0$, no change $= 1$, moderate improvement $= 2$, and substantial improvement $= 3$). The three raters agreed on 17/20

graphs, which is an agreement of 85%. On the three graphs where disagreement was reported, the rating that two observers agreed on was given. As previously mentioned, each graph was given a rating of 0, 1, 2 or 3 depending on the category it was allocated.

*Results of Ratings of Teacher Collected Data.* The improvements made for each child is provided according to the targeted behaviours was determined by the ratings of the teacher collected data. Of the 20 graphs of child behaviour change the following distribution of ratings resulted. Nine graphs were given the rating of 'substantial improvement'; four graphs were given the rating of 'moderate improvement'; and seven were given the rating of 'no change'. No child's behaviour was considered to deteriorate from baseline.

Ten teachers also completed a 6-week follow-up. Of the children who showed improvement following treatment ($n = 8$), seven demonstrated a maintenance of the change at 6-week follow-up. One child deteriorated at follow-up to baseline levels. The other two children, who did not show improvement following implementation of the intervention, had follow-up levels of behaviour that were consistent with baseline levels.

## Goal Achievement Scale

Of the 20 teachers involved in this study, six teachers achieved Goal Achievement Scale scores of 100%. The mean Goal Achievement Scale score was 69.9% (range of 0–100%).

## Sutter–Eyberg Student Behaviour Inventory and Teacher Report Form

The mean SESBI Intensity score at baseline was 116 and at post-intervention the mean was 95.0. In all except two cases the teacher reported decreases in the child's Sutter–Eyberg Student Behaviour Inventory Intensity Scores from pre- to post-treatment. Similarly, all except three teachers reported decreases in the number of problems from pre to post treatment (baseline mean = 10.3, post-treatment mean = 7.2).

The mean TRF Total Problem score at baseline was 60.5, and at post-treatment the mean was 57.2. Seven teachers did not report a decrease in the TRF Total Problem scores following intervention. On the Externalising Scale the baseline mean was 60.8 and the post-treatment mean was 58.3. Ten teachers did not report a decrease in the TRF Externalising Problem scores following intervention. There were some decreases in the TRF scores as a whole, however, they do not appear as marked as the SESBI scores.

A multivariate analysis of variance was performed with time as the factor and four dependent variables; SESBI Intensity, SESBI Problem, TRF Total and TRF Externalising. A significant multivariate effect was found, Wilks $\lambda = 0.44$, $F(4, 16) = 5.01$, $P = 0.008$. Subsequent univariate ANOVAs were conducted and significant differences were found between the groups on all four dependent variables: SESBI Intensity, $E(1, 19) = 15.77$, $P = 0.001$; SESBI Problem, $F(1, 19) = 12.68$, $P = 0.002$; TRF Total, $F(1, 19) = 7.58$, $P = 0.013$; TRF Externalising, $F(1, 19) = 6.52$, $P = 0.019$. All measures showed a significant decrease from pre to post intervention. The effect sizes were larger for the two SESBI measures than for the TRF measures.

TABLE I. Correlations between the Goal Achievement Scale, Treatment Integrity, SESBI and TRF

|  | Pretest SESBI | Pretest TRF | Treatment Integrity | Ratings of Graphs | GAS |
|---|---|---|---|---|---|
| **Pretest SESBI** | 1.0 | 0.66** | − 0.12 | − 0.48* | − 0.32 |
| **Pretest TRF** |  | 1.0 | 0.01 | − 0.48* | − 0.30 |
| **Integrity** |  |  | 1.0 | 0.60** | 0.67** |
| **Ratings of Graphs** |  |  |  | 1.0 | 0.80** |
| **GAS** |  |  |  |  | 1.0 |

\* Correlation is significant at the 0.05 level.
\*\* Correlation is significant at the 0.01 level.

*Consumer Satisfaction*

The consumer satisfaction ratings were high (M = 28.2, SD = 2.5), even for teachers whose target child's behaviour did not demonstrate improvement according to the daily record of behaviour, with 18 of the 20 teachers reporting consumer satisfaction levels of 80% or higher (i.e. greater than 25/30). The qualitative comments made by the teachers also clearly reflected their satisfaction with the teacher tip sheets. Some of the teachers commented that there were desirable side effects such as generalisation to other children, increased child self-esteem, generalisation across behaviours and generalisation of teacher behaviour. Only one teacher reported an unwanted side effect of the calling out tip sheet. This teacher said that the child became aggressive and answered back when the strategies were introduced. It was interesting to note that this child's behaviour improved, and the GAS score for that behaviour was 80.8 Indicating success with the calling out. The teacher did not provide any information on how frequently the answering back occurred, or whether it was an isolated incident.

*Relationship between Treatment Integrity, Child Behaviour Ratings and Outcome*

The relationships between treatment integrity, pretest rating of child behaviour (SESBI and TRF) and outcome (GAS and graphs of teacher collected data) were examined using Pearson product-moment correlations. These are summarised in Table I.

Using the ratings of graphs as the dependent variable, a regression analysis was conducted using pretest SESBI scores, pretest TRF and Treatment Integrity as the predictor variables. Treatment integrity significantly predicted the ratings of graphs, $F(1, 18 = 9.95, P = 0.005$. Treatment integrity accounted for 35.6% of variance in the ratings of graphs. SESBI scores added significantly to the prediction of ratings of graphs, $F(1, 18 = 5.25, P = 0.034$, accounting for 22.6% of variance in the ratings of graphs. This model accounted for 52.6% of the variance in ratings of graphs.

Using GAS scores as the dependent variable it was found that treatment integrity was a significant predictor of outcome, $F(1,18) = 14.36$, $P = 0.001$. Treatment integrity accounted for 44.9% of the variance in the GAS scores. Neither the SESBI nor the TRF added to the predicted variance.

**Discussion**

The first research question was whether child behaviour would improve following the use of the teacher tip sheets. The results of this study clearly indicate that most

children's behaviour did improve (on all measures) following intervention. The second research question was whether treatment integrity would be related to child behaviour outcome. It was found that treatment integrity was significantly related to child outcome. The closer the teacher adhered to the treatment programme, the better the child's outcome. The third research question was whether teachers would find the tip sheets acceptable. The consumer satisfaction results clearly indicate the tip sheets were highly acceptable to the teachers. Each of these findings will now be examined in detail.

### Child Outcome

The results indicate that the teacher tip sheets can be effective in improving discrete behaviour problems in the classroom setting. Of the 20 teachers involved in this study, 13 demonstrated the ability to improve child behaviour by implementing the strategies in the teacher tip sheets. Of the seven teachers whose children did not improve, or showed little improvement (as measured by ratings of the graphs of teacher collected data), three had low treatment integrity and the other four had target children whose behaviour fell into the borderline/clinical range on the pre-intervention measures of child behaviour.

In addition, the changes in child behaviour were evident across all measures of child behaviour, including the teacher report forms (SESBI and TRF). Past research has found that studies using teacher ratings of child behaviour were less likely to demonstrate reductions in the behaviours compared to the studies that used behavioural observation (Stage & Quiroz, 1997). Given this finding, the changes of behaviour effected in this study would appear to be quite pronounced.

The results of this study are consistent with the research on similar levels of parent intervention. A recent investigation by Sultana *et al.* (2000) found that parents using tip sheets for commonly occurring child behaviour problems reported improvements in their children's behaviour following intervention.

### Treatment Integrity

From the results of the multiple regression it was found that treatment integrity, on its own, was the most significant predictor of outcome (both GAS scores and ratings of graphs). Therefore, in regards to the second research question, it was found that the teachers' adherence to the treatment protocol was associated with child outcome. Therefore, the hypothesis that improvements in child behaviour would be related to treatment integrity was supported. This indicates that it is vital for the behavioural consultant to monitor the implementation of treatment, rather than just rely on the teachers' assurances that they are following the treatment regime. Anecdotally, all teachers reported to the consultant that they were implementing the tip sheet strategies as prescribed. However, the precise measurement of treatment integrity clearly indicated that this was not the case for all teachers.

Treatment integrity is an important measure in the behavioural consultation process, as it was significantly related to outcome. Prior to this study, Watson and colleagues had reported that 'there are no data directly linking integrity of the consultation process with direct measures of changes in either consultee's or client's behavior' (Watson *et al.*, 1997, p. 469). However, the current study has addressed this issue and the results clearly suggest that there is a relationship between treatment integrity and outcome. No

teacher who had low integrity ratings had a child whose target behaviour improved substantially.

Treatment integrity has been found to improve when performance feedback is provided to the teacher (Jones *et al.*, 1997). In the current study, those teachers with low integrity scores were not given feedback about their own behaviours. This would clearly be the next step in assisting these teachers, and would constitute a higher level intervention that is more time intensive and requires more frequent psychologist contact.

Treatment Integrity was not the only factor in predicting child outcome. The child's level of behaviour problems prior to intervention was also a factor in predicting outcome. This will now be discussed.

### Relationship Between Pretest Ratings of Child Behaviour and Outcome

Although treatment integrity was the most significant predictor of outcome success, the pretest ratings of child behaviour were also found to be associated with child outcome.

The finding that lack of improvement in behaviour problems (as measured by ratings of the graphs) was significantly related to scores in the borderline/clinical range on a measure of global child behaviour before intervention provides strong support for the need for a hierarchical model in education.

In a study by Little *et al.* (2000), it was found that children who had more severe behaviour problems were found to be more likely to demonstrate behaviour problems across settings. In the current study, it was found that these children were less likely to respond to minimal interventions than children with more moderate behaviour problems. Those children with more severe behaviour problems may require a higher level intervention than the tip sheets.

It is important to note that falling into the borderline/clinical range on measures of child behaviour did not automatically result in a failure for the behaviour problem to improve. In fact, three of the children who did demonstrate substantial behaviour change had pretest scores in the borderline/clinical range on the TRF or SESBI. However, of the children who did fail to improve, all except one had clinically significant general behaviour problems on at least one of the measures of child behaviour. This indicates that some, but not all, children with high levels of behaviour problems will respond to minimal interventions such as the teacher tip sheets.

It would be of interest to investigate whether there was some systematic difference between the children with clinically significant general behaviour problems who improved as compared to those who did not improve. Perhaps, those who improved were only demonstrating the behaviour problems in the one setting, and those that did not improve had behaviour problems in multiple settings.

### Relationship Between Measures of Child Outcome

The Goal Achievement Scale is currently not a widely used measure of treatment outcome, and the validity of it as an appropriate measure of treatment outcome could be questioned. In the current study, a correlation of .8 was found between the GAS scores and the ratings of the graphs. The GAS is a very simple measure in that it only requires the initial baseline data to be considered, and then the teacher and researcher agree on a level that would constitute 100% success. In contrast, the use of graphed behaviour outcomes requires the researcher to firstly graph all intervention data, then

obtain at least two independent ratings of the graphs, and finally compare these ratings to determine a final rating. In comparison to the GAS, this is quite a time consuming process. Hence, future studies could use the GAS score as a simple and valid measure of treatment outcome.

## Generalisation of Treatment Effects

While generalisation of treatment effects across children, behaviours and settings was not directly measured, there was evidence to suggest generalisation did occur with some of the sample. Anecdotal reports from the teachers indicated that the treatment effects often generalised to other children. The results of the SESBI measures (intensity and frequency) also suggest that there may have been some generalisation across behaviour. Of the 20 teachers who completed the SESBI, 17 reported decreases in both the intensity and frequency ratings. Given that the SESBI lists 36 behaviours, including the target behaviours, a decrease in the overall score may indicate improvement in the child's general behaviour as well as the target behaviour. For at least one child, there was anecdotal support for generalisation across settings. Three teachers reported that either other teachers or parents had noticed improvements in the child's behaviour in other situations. Seven of the eight teachers who demonstrated behavioural improvement in their child following intervention, had maintained these improvements at follow-up. Only one child's behaviour deteriorated to baseline levels. Therefore, maintenance appears to have occurred for the sample.

## Examination of Teacher Characteristics and Drop Out Rates

As only one of 21 teachers withdrew from the study, it is of interest to consider the reasons for this. The teacher did not complete the pretest questionnaires or the daily recording of behaviour because it was 'too time consuming' for her. Anecdotally, the teacher appeared quite stressed and stated a number of times that the requirements for the study were too much, and she could not understand how teachers were able to complete all that was required. The teacher had forgotten both appointments that were made with her to discuss the project, and appeared to be quite disorganised with her lesson preparation. On both occasions she stated that she did not have time to talk because she was not sure what she was doing for the day and needed to plan her lessons.

In terms of a hierarchical model of school intervention, it appears that this teacher might be a good candidate for a higher level intervention that addressed issues such as time management, stress management, as well as general classroom management. A minimal level intervention may have been appropriate to manage the child's behaviour: however, it is clear that the teacher needed assistance with managing her own behaviour (stress and time). While it was beyond the scope of the current research paradigm to develop and investigate the higher level interventions, this is clearly an area that warrants further research.

## Consumer Satisfaction

The hypothesis that the teacher tip sheets would be highly acceptable to teachers was supported by the results of this study. The consumer satisfaction measures, which

included questions on acceptability, indicated that teachers approved of the strategies used and perceived them to be effective in managing children's behaviour problems. This is consistent with previous research by the authors, which found high levels of acceptability for the teacher tip sheets. It is of interest that even teachers who did not have high levels of success in reducing the disruptive behaviour of their children (as evident from the graphs of teacher collected data), still reported that the treatment was acceptable and were satisfied with the treatment approach. Some of the teachers even reported improvements in the child's behaviour despite the evidence from the teacher collected data indicating no change. This highlights the need for multiple outcome measures, as the consumer satisfaction measure alone may not be indicative of the true levels of behaviour change.

## Conclusions

The teacher tip sheets are clearly an effective minimal intervention to deal with behaviour problems in school age children. This type of intervention would fit into a hierarchical model for school intervention such as that developed by Little & Hudson (1998). Given that this minimal level intervention was not effective for all students and teachers, a higher level of intervention may have been required to produce behavioural improvements in those students.

Further study is now needed to examine the use of these strategies, or the development of alternative strategies, for children at high school level. Also, an examination is needed of whether teachers who are nominated as needing assistance in classroom management have similar results as the teachers who volunteered for this study.

*Correspondence:* Dr E. Little, Department of Psychology and Intellectual Disability Studies, RMIT University, Bundoora 3083, Australia (e-mail: emma.little@ rmit.edu.au).

## REFERENCES

Achenbach, T.M. & Edelbrock, C.S. (1983). *Manual for the Child Behavior Checklist and Child Behavior Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T.M. & Edelbrock, C.S. (1986). *Manual for the Teacher's Report Form and Teacher Version of the Child Behavior Profile*, revised. Burlington, VT: Department of Psychiatry, University of Vermont.

Armstrong, K.J., Ehrhardt, K.E., Cool, R.T. & Poling, A. (1997). Social validity and treatment integrity data: reporting in articles published in journal of developmental and physical disabilities, 1992–1995. *Journal of Developmental and Physical Disabilities*, *9*, 359–367.

Australian Bureau of Statistics (1992). *1991 Census Socio-economic Indexes for Areas*. Canberra: Alan Law.

Department of Human Services (1996). *Parenting Tip Sheets*. Victoria: Primary Care Division, Victorian Government Department of Human Services.

Eme, R.F. & Kavanaugh, L. (1995). Sex differences in conduct disorder. *Journal of Clinical Child Psychology*, *24*, 406–426.

Eyberg, S. (1974). *Therapy Attitude Inventory*. FL: Author.

Eyberg, S. (1992). Parent and teacher behavior inventories for the assessment of conduct problem behaviors in children. In: L. VandeCreek, S. Knapp, & T.L. Jackson (Eds), *Innovations in Clinical Practice: A source book*, Vol. 11 (pp. 261–266). Sarasota, FL: Professional Resource Press.

Funderburk, B.W. & Eyberg, S.M. (1989). Psychometric characteristics of the Sutter–Eyberg Student behavior Inventory: A school behavior rating scale for use with preschool children. *Behavioral Assessment*, *11*, 297–313.

Hogue, A., Liddle, H.A. & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy*, *33*, 332–345.

Hops, H. & Cobbs, J.A. (1974). Initial investigations into academic survival-skill training, direct instruction and first grade achievement. *Journal of Educational Psychology*, *66*, 548–553.

Hudson, A. (1998). Evaluation of clinical interventions using the goal achievement scale, paper presented at the 21st National Conference of the Australian Association for Cognitive and Behaviour Therapy, Adelaide.

Hudson, A., Jauernig, R., Wilken, P. & Radler, G. (1995). Regionally based teams for the treatment of challenging behaviour: A 3 year outcome study. *Behaviour Change*, *12*, 209–211.

Jones, K.M., Wickstrom, K.F. & Friman, P.C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly*, *12*, 316–326.

Kazdin, A.E. (1989). Conduct disorder. *Psychiatric Hospital*, *20*, 153–158.

Kazdin, A.E., Bass, D., Ayers, W.A. & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology*, *58*, 729–740.

Keller, H.R. & Tapasak, R.C. (1997). Classroom management. In: A. P. Goldstein and J. Close (Eds), *School Violence Intervention: A practical handbook* (pp. 107–126). New York, NY: Guilford Press.

Little, E. & Hudson, A. (1998). Conduct problems and treatment across home and school: A review of the literature. *Behaviour Change*, *15*, 213–227.

Little, E., Hudson, A. & Wilks, R. (2000). Conduct problems across home and school. *Behaviour Change*, *17*, 1–9.

Loeber, R., Stouthamer-Loeber, M. & Green, S.M. (1991). Age at onset of problem behaviour in boys, and later disruptive and delinquent behaviours. *Criminal Behaviour and Mental Health*, *1*, 229–246.

McConaughy, S.H., Achenbach, T.M. & Gent, C.L. (1988). Multiaxial empirically based assessment: Parent, teacher, observational, cognitive and personality correlates of child behavior profile types for 6- to 11-year old boys. *Journal of Abnormal Child Psychology*, *16*, 485–509.

Robins, L.N. (1991). Conduct disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *32*, 193–209.

Sanders, M.R. (1999). *The Triple-P Positive Parenting Program: Towards an empirically validated multi-level parenting and family support strategy for the prevention and treatment of child behavior and emotional problems. Clinical Child & Family Psychology Review*, *2*, 71–90.

Sanders, M.R. & Dadds, M.R. (1993). *Behavioral Family Intervention*. Needham Heights, MA: Allyn and Bacon.

Schaughency, E.A., Hurley, L.K., Yano, K.E., Seeley, J.R. & Talarico, B. (1989). Psychometric properties of the Sutter–Eyberg Student Behavior Inventory with Clinic Referred Children, paper presented at the annual meeting of the American Psychological Association, New Orleans, LA.

Shinn, M.R., Ramsey, E., Walker, H.M., O'Neill, R.E. & Steiber, S. (1987). Antisocial behavior in school settings: Initial differences in an at-risk and normal population. *Journal of Special Education*, *21*, 69–84.

Stage, S.A. & Quiroz, D.R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review*, *26*, 333–368.

Sultana, C., Matthews, J., DeBortoli, D. & Cann, W. (2000). *An evaluation of two levels of the positive parenting program (TRIPLE P) delivered by primary care practitioners*. Unpublished manuscript.

Sutter, J. & Eyberg, S. (1984). *Sutter–Eyberg Student Behavior Inventory*. Gainesville, FL: Author.

Tarnowski, K.J. & Simonian, S.J. (1992). Assessing treatment acceptance: The Abbreviated Acceptability Rating Profile. *Journal of Behaviour Therapy and Experimental Psychiatry*, *23*, 101–106.

Watson, T.S., Sterling, H.E. & McDade, A. (1997). Demythifying behavioral consultation. *School Psychology Review*, *26*, 467–474.

Wheldall, K. & Merrett, F. (1988). Which classroom behaviours do primary school teachers say they found the most troublesome? *Educational Review*, *40*, 13–27.